

# Systematic misperceptions of 3-D motion explained by Bayesian inference

**Bas Rokers**

Department of Psychology, University of Wisconsin,  
Madison, WI, USA



**Jacqueline M. Fulvio**

Department of Psychology, University of Wisconsin,  
Madison, WI, USA



**Jonathan W. Pillow**

Department of Psychology, University of Wisconsin,  
Madison, WI, USA



**Emily A. Cooper**

Department of Psychology, University of Wisconsin,  
Madison, WI, USA



People make surprising but reliable perceptual errors. Here, we provide a unified explanation for systematic errors in the perception of three-dimensional (3-D) motion. To do so, we characterized the binocular retinal motion signals produced by objects moving through arbitrary locations in 3-D. Next, we developed a Bayesian model, treating 3-D motion perception as optimal inference given sensory noise in the measurement of retinal motion. The model predicts a set of systematic perceptual errors, which depend on stimulus distance, contrast, and eccentricity. We then used a virtual-reality headset as well as a standard 3-D desktop stereoscopic display to test these predictions in a series of perceptual experiments. As predicted, we found evidence that errors in 3-D motion perception depend on the contrast, viewing distance, and eccentricity of a stimulus. These errors include a lateral bias in perceived motion direction and a surprising tendency to misreport approaching motion as receding and vice versa. In sum, we present a Bayesian model that provides a parsimonious account for a range of systematic misperceptions of motion in naturalistic environments.

## Introduction

The accurate perception of visual motion is critical for everyday behavior. In the natural environment, motion perception involves determining the 3-D direction and speed of moving objects, based on both retinal and extraretinal sensory cues. In the laboratory, a large number of studies have reported systematic biases in the perception of 3-D motion, despite the availability

of many such cues (Fulvio, Rosen, & Rokers, 2015; Harris & Dean, 2003; Harris & Drga, 2005; Lages, 2006; Rushton & Duke, 2007; Welchman, Lam, & Bülhoff, 2008; Welchman, Tuck, & Harris, 2004). These perceptual errors may contribute to behavioral failures in real-world scenarios, such as catching projectiles (Peper, Bootsma, Mestre, & Bakker, 1994) and driving under foggy conditions (Pretto, Bresciani, Rainer, & Bülhoff, 2012; Shrivastava, Hayhoe, Pelz, & Mruczek, 2010; Snowden, Stimpson, & Ruddle, 1998). Here we ask if a range of systematic errors in 3-D motion perception can be understood as a consequence of 3-D viewing geometry and reasonable prior expectations about the world.

Bayesian-observer models are a strong candidate for addressing this question. They provide a straightforward rule for the optimal combination of incoming sensory evidence with prior knowledge. The Bayesian framework has successfully explained a variety of perceptual phenomena (Girshick, Landy, & Simoncelli, 2011; Knill, 2007; Knill & Richards, 1996), including systematic biases in 2-D motion perception (Weiss, Simoncelli, & Adelson, 2002). Specifically, when visual input is unreliable (for example, when a stimulus has low contrast), observers systematically underestimate the speed of visual motion in the fronto-parallel plane: Low-contrast patterns appear to move more slowly than otherwise equivalent high-contrast patterns (Stone & Thompson, 1992; Thompson, 1982). This misperception, along with several other seemingly unrelated phenomena in motion perception, can be elegantly accounted for by a Bayesian model that incorporates a prior assumption that objects in the world tend to move

Citation: 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/jov.18.3.23>.

<https://doi.org/10.1167/jov.18.3.23>

Received June 7, 2017; published March 29, 2018

ISSN 1534-7362 Copyright 2018 The Authors



slowly (Hürlimann, Kiper & Carandini, 2002; Stocker & Simoncelli, 2006; Weiss et al., 2002).

In the 3-D domain, a prior assumption for slow motion may have additional consequences. For example, observers exhibit a *lateral bias*: They systematically overestimate angle of approach in 3-D, such that objects moving toward the head are perceived as moving along a path that is more lateral than the true trajectory (Harris & Dean, 2003; Harris & Drga, 2005; Lages, 2006; Rushton & Duke, 2007; Welchman et al., 2004; Welchman et al., 2008). Bayesian models of 3-D motion perception, assuming a slow motion prior, can account for this bias (Lages, 2006; Lages, Heron, & Wang, 2013; Wang, Heron, Moreland, & Lages, 2012; Welchman et al., 2008). However, existing models are restricted to specific viewing situations (stimuli in the midsagittal plane) and have been tested using tasks and stimuli that limit the kind of perceptual errors that can be observed. In addition, these models have not addressed a recently identified perceptual phenomenon in which the direction of motion in depth (but not lateral motion) is frequently misreported: Approaching motion is reported to be receding and vice versa (Fulvio et al., 2015).

Here, we provide a Bayesian model of 3-D motion perception for arbitrary stimulus locations and naturalistic tasks. The derived model provides predictions for both average biases and trial-to-trial variability. We follow up this model with a series of perceptual studies, both in virtual reality and on a stereoscopic desktop display, and find that the model predicts the impact of stimulus distance, contrast, and eccentricity on the magnitude of 3-D motion misperceptions. We thus provide a unified account of multiple perceptual phenomena in 3-D motion perception, showing that geometric considerations, combined with optimal inference under sensory uncertainty, explain these systematic and, at times, dramatic misperceptions.

## Developing a Bayesian model

### Geometric explanation of biases in 3-D motion perception

The Bayesian-brain hypothesis posits that perception of the physical world is dictated by a probabilistic process that relies on a distribution called the posterior (Kersten & Yuille, 2003; Knill & Pouget, 2004; Knill & Richards, 1996). This posterior  $P(s|r)$  specifies the conditional probability of the physical stimulus  $s$  given a sensory measurement or response  $r$ . The posterior is determined, according to Bayes's rule, by the product of two probabilistic quantities known as the likelihood and the prior. The likelihood  $P(r|s)$  is

the conditional probability of the observed sensory response  $r$  given a physical stimulus  $s$ . It characterizes the information that neural responses carry about the sensory stimulus. Increased sensory uncertainty, due to ambiguity or noise in the external world or internal noise in the sensory system, manifests as an increase in the width of the likelihood. The prior  $P(s)$  represents the observer's assumed probability distribution of the stimulus in the world. The prior may be based on evolutionary or experience-based learning mechanisms. The relationship between posterior, likelihood, and prior is given by Bayes's rule, which states:

$$P(s|r) \propto P(r|s)P(s). \quad (1)$$

When sensory uncertainty is high, the likelihood is broad and the prior exerts a relatively large influence on the posterior, resulting in percepts that are systematically more biased toward the prior (but see Wei & Stocker, 2015). That is, the visual system relies on prior assumptions when sensory information is unreliable. Misperceptions will inevitably occur when actual stimulus properties diverge from these prior assumptions, particularly when sensory uncertainty is high.

Here we apply this Bayesian framework to the problem of 3-D motion perception. Since the derivation of the posterior distribution for 3-D motion is lengthy, we first provide an intuition by examining a simple diagram illustrating the consequences of perspective projection on retinal signals to 3-D motion. First, we consider that light reflected from moving objects in the world will project through the optics of the eye and cast a pattern with a particular angular velocity on the retina. This is illustrated in Figure 1A. A simplified top-down diagram illustrates the left and right eyes of an observer (projections are shown for the left eye only; the right eye is for reference). Two linear motion vectors are illustrated in orange and green. The vectors have the same length, indicating the same speed in the world, but they move in different directions, either in depth (toward the observer, shown in green) or laterally (to the left, shown in orange). Of course, the angular-velocity signal in either eye alone does not specify the direction of motion in the world. While these signals do constrain the possible trajectories, estimates of 3-D motion critically depend on the *relationship* of the retinal-velocity signals between the two eyes.

The angular subtense of each vector in the left eye is illustrated by the green and orange arcs, respectively. Note that although the vectors have the same length, and thus the same world speed, the angular subtense of the vector corresponding to motion in depth is considerably smaller than the one corresponding to lateral motion, and thus produces a considerably slower retinal speed.

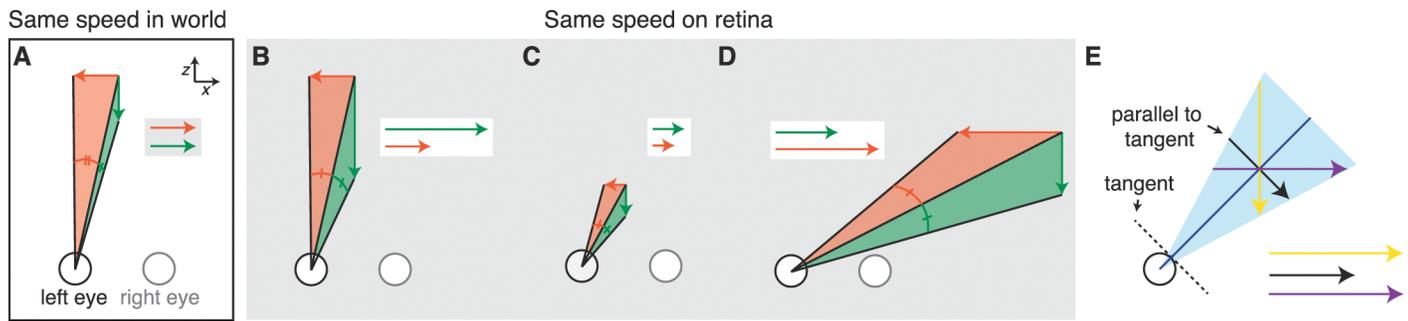


Figure 1 Schematic top-down view illustrating how uncertainty in retinal velocity propagates asymmetrically to motion trajectories in the world. (A) Two orthogonal motion vectors with the same speed in the world (motion in depth in green and lateral motion in orange) project to different angular speeds on the retina. (B) A fixed retinal speed projects to a longer vector for motion in depth than for lateral motion. The same geometry applies to the transformation of uncertainty. (C) This difference is much reduced at near viewing distances. (D) This relationship can invert for trajectories that occur off of the midsagittal plane. (E) Illustration of how the tangent line of a circle determines the vector direction with the minimum length for a given angle and distance. Note that when motion is directly toward either eye, this will project to zero retinal velocity in one eye (ignoring looming/optical expansion) and nonzero velocity in the other.

Next, we consider that our perception of motion in the world (i.e., the motion vectors) relies on measuring these angular speeds (i.e., the arcs) and inferring the physical motion trajectories that caused them. To examine how the limited accuracy in measuring angular speed on the retina propagates to limited accuracy in perceiving different motion trajectories, we can project a fixed retinal uncertainty in angular speed back into world coordinates. This is illustrated in Figure 1B, again just for two sample directions of motion directly toward or to the left of the observer. Note that even though the angular speeds on the retina are the same, the amount of uncertainty for motion in depth (represented by the vector length) is greater than for lateral motion. That is, a given uncertainty in angular speed will result in greater uncertainty for motion in depth in the world. Vectors are reproduced side by side on the right for clarity. This difference is simply due to inverting the projection shown in Figure 1A. Note that this is an observation about the geometry of 3-D viewing, not a claim about how observers estimate motion direction.

However, is the high uncertainty for motion in depth universally true for all viewing situations? Simple geometric diagrams show that this is not the case. Figure 1C and 1D illustrates two additional situations. In Figure 1C, the distance of the motion vectors from the eyes is decreased. Uncertainty is still larger for motion in depth, but the increase relative to lateral motion is substantially attenuated. In Figure 1D, the motion vectors are located off to the observer's right. In this case, the relationship has actually inverted, and uncertainty for lateral motion is greater. Note that we only illustrate motion directly lateral to or toward the observer. However, as we will show later, since any motion vector can be decomposed into its components

along these orthogonal axes, these general principles will hold for any direction.

Indeed, if we model the eye as a circle and assume the center of projection is at the circle's center, it is easy to see that there is no consistent increase in uncertainty for motion in depth relative to lateral motion. Intuitively, for a given trajectory the motion component parallel to the tangent of the circle (at the point where a line connecting the center of projection to the vector intersects the circle) will have the least uncertainty (Figure 1E). In the derivation that follows, we quantitatively determine the predicted uncertainty for 3-D motion trajectories in all directions under any viewing situation and use these predictions to formulate a Bayesian model for 3-D motion perception, focusing specifically on the impact of uncertainty in retinal motion on errors in perceived direction of objects moving through the world.

## Relationship between 3-D motion trajectories and retinal velocities

We can describe the motion of any object in space relative to an observer in a 3-D coordinate system with the function

$$p(t) = [x(t), y(t), z(t)], \quad (2)$$

where  $p$  is the position of the object as a function of time  $t$  in a coordinate system defined over  $x$ -,  $y$ -, and  $z$ -axes. Here, we use a head-centered coordinate system and place the origin at the midpoint between the two eyes of the observer (see icon in upper left corner of Figure 2). In this left-handed coordinate system,

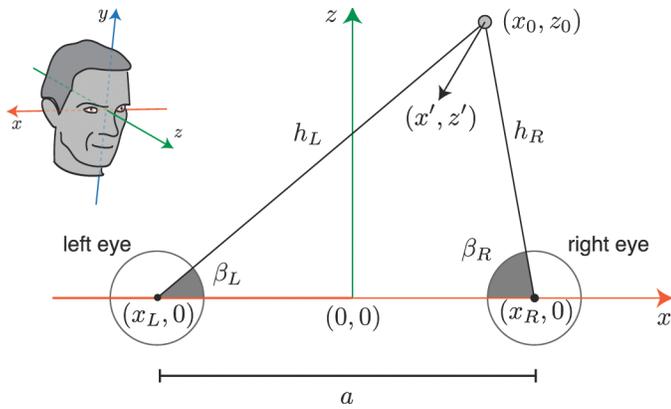


Figure 2 Diagram of the 3-D motion coordinate system. The icon in the upper left shows the origin and axes of the coordinate system, with arrowheads indicating the positive direction on each axis. The top-down view shows a slice through the interocular axis in the  $xz$ -plane. Large circles indicate the left and right eyes. The smaller gray circle and arrow indicate the location and trajectory of motion of an object. The coordinates of key points are indicated in  $x$  and  $z$  ( $y = 0$  for all points), as well as several line segments and angles. Note that  $x_0$  and  $z_0$  denote the coordinates of the object with the motion defined by Equation 2, evaluated at time point  $t = t_0$ .

the  $x$ -axis is parallel to the interocular axis (positive rightward), the  $y$ -axis is orthogonal to the  $x$ -axis in the plane of the forehead (positive upward), and the  $z$ -axis extends in front of and behind the observer (positive in front).

We will model the retinal information available from horizontal velocities, and thus consider the projection of points onto the  $xz$ -plane ( $y = 0$  for all points; Figure 2). Note, however, that this does not mean that this model is valid only for stimuli in the plane of the interocular axis. As long as retinal angles are represented in an azimuth-longitude coordinate system, the horizontal retinal velocities can be computed from the  $x$  and  $z$  components of 3-D motion vectors alone. This geometry is independent of the observer’s point of fixation but assumes that fixation does not change over the course of stimulus presentation. In this coordinate system, the  $(x, z)$  coordinates of the left and right eye are defined as  $(x_L, 0)$  and  $(x_R, 0)$ , respectively. The distance between the eyes along the interocular axis, denoted by  $a$ , is  $x_R - x_L$ .

At any time point, an object with coordinates  $(x(t), z(t))$  will project to a different horizontal angle in each eye. If we define these angles relative to the  $x$ -axis in the  $xz$ -plane, they are given by

$$\beta_{L,R}(t) = \arctan\left(\frac{z(t)}{x(t) - x_{L,R}}\right), \quad (3)$$

where  $\beta_L(t)$  and  $\beta_R(t)$  indicate the angle in the left and the right eye, respectively. The object will generally have

a different distance from each eye. These distances are given by

$$h_{L,R}(t) = \sqrt{(x(t) - x_{L,R})^2 + z(t)^2}, \quad (4)$$

where  $h_L(t)$  and  $h_R(t)$  indicate the distance from the left and the right eye, respectively.

Since we are interested in motion cues, we differentiate Equation 3 with respect to time to determine the relationship between object motion and motion on the retina. We denote first derivatives of functions with the convention  $df(x)/dt = f'(x)$ . This yields

$$\beta'_{L,R}(t) = \frac{1}{1 + \left(\frac{z(t)}{x(t) - x_{L,R}}\right)^2} \left[ \frac{z'(t)(x(t) - x_{L,R}) - z(t)x'(t)}{(x(t) - x_{L,R})^2} \right]. \quad (5)$$

Rearranging Equation 5 and substituting in Equation 4 allows us to simplify to

$$\beta'_{L,R}(t) = \frac{1}{h_{L,R}(t)^2} [z'(t)(x(t) - x_{L,R}) - z(t)x'(t)]. \quad (6)$$

For motion estimation,  $\beta'_L(t)$  and  $\beta'_R(t)$  are the sensory signals, representing retinal velocities in left and right eyes, and the motion components in the world  $x'(t)$  and  $z'(t)$  that generated them are unknown. We therefore solve for  $x'(t)$  and  $z'(t)$  as a function of  $\beta'_L(t)$  and  $\beta'_R(t)$ . For simplicity, we will drop the time index so that  $\beta'_{L,R}$ ,  $h_{L,R}$ ,  $z_0$ ,  $z'$ ,  $x_0$ , and  $x'$  refer to  $\beta'_{L,R}(t)$ ,  $h_{L,R}(t)$ ,  $z(t)$ ,  $z'(t)$ ,  $x(t)$ , and  $x'(t)$ , each evaluated at time  $t = t_0$ . To determine the velocity  $x'$  in terms of retinal velocities, we rearrange Equation 6 for the left eye to solve for  $z'$ , substitute the result back into Equation 6 for the right eye, and solve for  $x'$ , yielding

$$x' = \frac{1}{z_0 a} [\beta'_L h_L^2 (x_0 - x_R) - \beta'_R h_R^2 (x_0 - x_L)]. \quad (7)$$

Recall that  $a$  refers to the interocular separation. To determine the equation for  $z'$  in terms of retinal velocities, we rearrange Equation 6 for the left eye to solve for  $x'$  and substitute this back into Equation 6 for the right eye, yielding the following equation for  $z'$ , also in terms of retinal velocities:

$$z' = \frac{1}{a} [\beta'_L h_L^2 - \beta'_R h_R^2]. \quad (8)$$

### Propagating uncertainty for 3-D motion

We assume that the measurements of retinal motion in each eye,  $\beta'_L$  and  $\beta'_R$ , are corrupted by independent additive noise:

$$\ddot{\beta}'_L = \beta'_L + n_{\beta'_L} \quad (9A)$$

$$\ddot{\beta}'_R = \beta'_R + n_{\beta'_R} \quad (9B)$$

Here,  $\ddot{\beta}'_L$  and  $\ddot{\beta}'_R$  denote the measured retinal velocities in the left and right eye, respectively, where noise samples  $n_{\beta'_L}$  and  $n_{\beta'_R}$  are assumed to be independently drawn from a zero-mean Gaussian distribution with a variance of  $\sigma_{n_{\beta'}}$ . Note that the assumption of constant additive noise is inconsistent with Weber’s law (which would predict that the noise increases proportionately with speed), and that this model assumes that noise is independent of the motion direction. However, psychophysical experiments have shown that for relatively slow speeds (less than  $\sim 2^\circ/s$ – $4^\circ/s$ ), speed-discrimination thresholds are more stable than is predicted by Weber’s law (Freeman, Champion, & Warren, 2010; McKee, Silverman, & Nakayama, 1986; Stocker & Simoncelli, 2006).

We should note that although our derivations depend on the locations of the object relative to the eyes, they do not depend on *where* the observer fixates. This independence occurs because the position of fixation does not affect the angular velocity cast by a moving object at a given head-centric location—assuming, as in previous models, that fixation remains stable during stimulus presentation (Lages, 2006; Wang et al., 2012; Welchman et al., 2008). However, this does not explicitly account for any differences in retinal-velocity estimation across the visual field. For example, while retinal-motion signals may be less reliable at eccentricity compared to at fixation (Johnston & Wright, 1986; Levi, Klein, & Aitsebaomo, 1984), we do not explicitly incorporate such differences here.

Under the assumption that the object’s initial 3-D location—its distance  $z_0$  and its location relative to each eye ( $x_0 - x_L$ ) and ( $x_0 - x_R$ )—is known, we can use Equations 7 and 8 (which specify  $x'$  and  $z'$  as linear combinations of  $\beta'_L$  and  $\beta'_R$ ) to determine the noise covariance of the sensory measurements of speed in  $x$  and  $z$  ( $\ddot{x}'$  and  $\ddot{z}'$ ). First, we rewrite the linear transformation from retinal velocities  $\beta' = (\beta'_L, \beta'_R)$  to world velocities  $w' = (x', z')$  in terms of the matrix equation  $w' = A\beta'$ . In this formulation,  $A$  is given by

$$A = \begin{bmatrix} \frac{h_L^2(x_0 - x_R)}{z_0 a} & \frac{-h_R^2(x_0 - x_L)}{z_0 a} \\ \frac{h_L^2}{a} & \frac{-h_R^2}{a} \end{bmatrix}. \quad (10)$$

If we assume independent and equal noise distributions for the two eyes, the noise covariance of the sensory measurements is given by  $M = AA^T\sigma_{n_{\beta'}}^2$ , which is equal to

$$M = \text{cov}(\ddot{x}', \ddot{z}') = \begin{bmatrix} \frac{(x_0 - x_R)^2 h_L^4 + (x_0 - x_L)^2 h_R^4}{z_0^2 a^2} \sigma_{n_{\beta'}}^2 & \frac{(x_0 - x_R) h_L^4 + (x_0 - x_L) h_R^4}{z_0 a^2} \sigma_{n_{\beta'}}^2 \\ \frac{(x_0 - x_R) h_L^4 + (x_0 - x_L) h_R^4}{z_0 a^2} \sigma_{n_{\beta'}}^2 & \frac{h_L^4 + h_R^4}{a^2} \sigma_{n_{\beta'}}^2 \end{bmatrix}. \quad (11)$$

To gain more intuition for the relative noise effects on the  $x$  and  $z$  velocity components of a motion trajectory, we plot the sensory uncertainty for each velocity component (the square root of the diagonal elements of Equation 11, denoted  $\sigma_{x'}$  and  $\sigma_{z'}$ ) as a function of horizontal location  $x_0$  and distance in depth  $z_0$  in Figure 3A and 3B. Each panel contains an isocontour plot showing the log of the sensory uncertainty at each true spatial location. Several features are notable. The uncertainty in  $x'$  is at its minimum for points that fall on or near the midsagittal plane and increases for points to the left and right. The uncertainty in  $z'$  is at its minimum for points closest to the eyes and increases radially away from the midpoint between the eyes. Note that uncertainty in  $x'$  also increases with distance, but not as steeply as in  $z'$ . In the central visual field, the uncertainty in  $z'$  is generally much greater than the uncertainty in  $x'$ .

To illustrate the relative magnitude of uncertainty in  $x'$  and  $z'$ , we plot the log of the ratio of the two values for a subset of points close to the observer in Figure 3C (within 25 cm left/right and 100 cm in depth). Ratios greater than 0 (red) indicate that uncertainty in  $z'$  is greater than  $x'$ , and ratios less than 0 (blue) indicate the reverse. In the central visual field, this ratio is greater than 1. This is consistent with previous work (Welchman et al., 2008). However, the ratio varies considerably as a function of both viewing distance and viewing angle. At steep viewing angles ( $>45^\circ$ ), the relationship reverses and  $x'$  uncertainty is actually greater than  $z'$  uncertainty. We should note that our model includes uncertainty only in object speed, not in object location. Uncertainty in object location would likely increase for objects projecting to larger retinal eccentricities.

Equation 11 indicates that the uncertainties in  $x'$  and  $z'$  are not independent. To visualize this relationship, in Figure 3D we show the covariance ellipses for a set of locations within 100 cm in depth (the inset shows a zoomed view of nearby points). For most locations, the ellipses are highly elongated, indicating that for each location, uncertainty is anisotropic across directions. As expected from the geometric analysis (Figure 1), the axis of minimal uncertainty is orthogonal to a line connecting each location back to the interocular axis, independent of the direction of gaze. This creates

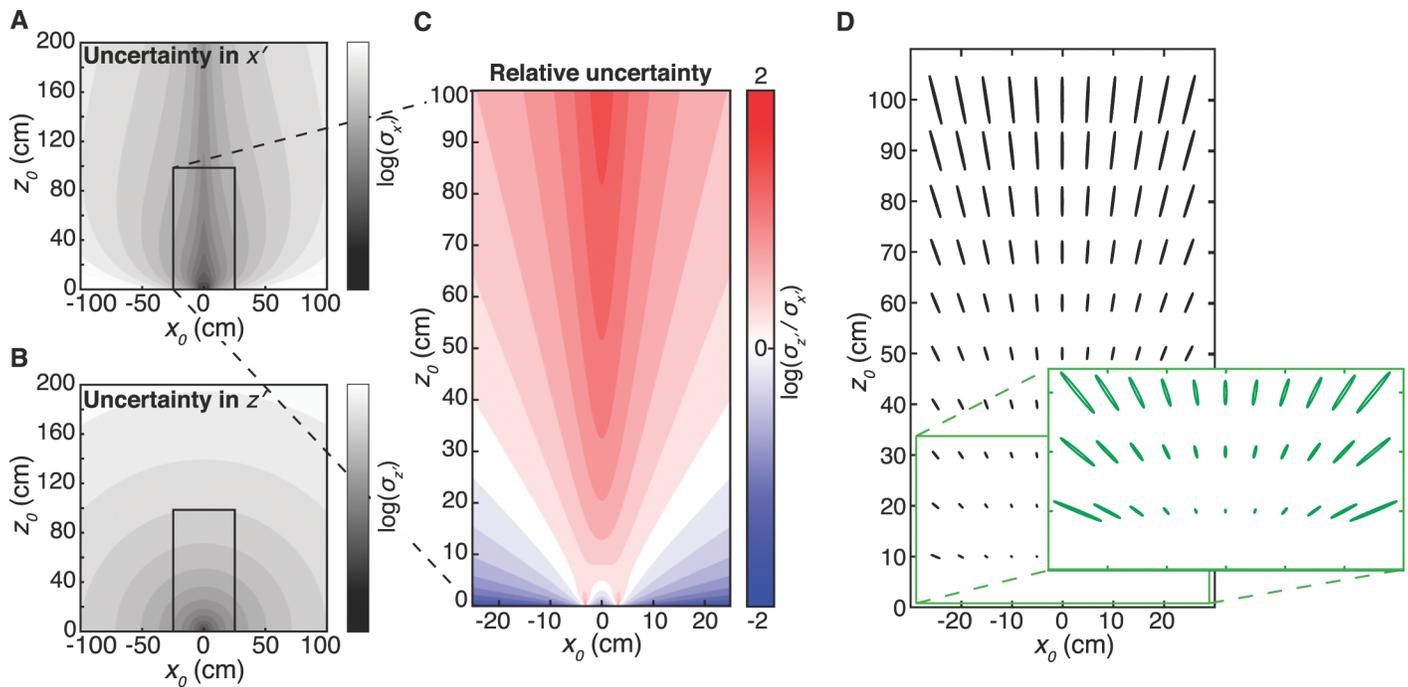


Figure 3 Uncertainty for  $x$  and  $z$  motion vary with stimulus distance and head-centric eccentricity. (A) Uncertainty in the  $x$  component of a motion vector ( $x'$ ) is plotted in arbitrary units as a function of location in  $x$  and  $z$  (assuming an interocular distance  $a = 6.4$  cm). (B) Same as (a), except for the  $z$  component of motion ( $z'$ ). The color-map scales of (A) and (B) are the same. (C) The ratio between the values in the boxed region in (A) and (B). (D) Ellipses illustrate the noise covariance of  $x'$  and  $z'$  for a range of spatial locations. Ellipse scale indicates the relative uncertainty for each location, and orientation indicates the axis of maximal uncertainty. All ellipses have been reduced by scaling with an arbitrary factor to fit within the plot. Inset shows the same ellipses for a small spatial region (also with a different scaling).

a radial pattern, in which uncertainty is highest for motion extending radially from the observer's location. Along the midsagittal plane ( $x_0 = 0$ ), the covariance is zero and the axes of minimal and maximal uncertainty align with the  $x$ - and  $z$ -axes, respectively.

Indeed, if we consider only cases in which the stimulus is presented in the midsagittal plane—as is often done in perceptual studies—the off-diagonal elements of the covariance matrix become zero, and we can simplify  $\sigma_{x'}^2$  and  $\sigma_{z'}^2$  to

$$\sigma_{x'}^2 = \frac{h^4}{2z_0^2} \sigma_{n_{p'}}^2 \quad (12)$$

$$\sigma_{z'}^2 = \frac{h^4}{2(a/2)^2} \sigma_{n_{p'}}^2, \quad (13)$$

where  $h = \sqrt{(a/2)^2 + z_0^2}$ . Typical viewing conditions (where  $z_0 \gg a$ ) result in substantially larger uncertainty for the  $z$  component of velocity than for the  $x$  component. However, if  $z_0$  is equal to  $a/2$ , half the interocular distance, the variances in  $x'$  and  $z'$  will be equal. Thus, while uncertainty for motion in depth in

the midsagittal plane clearly tends to be substantially higher than for lateral motion, the relative uncertainty is reduced for near viewing distances.

### Application of Bayes's rule to predict perceived motion in the midsagittal plane

In order to predict how sensory uncertainty in motion measurement affects actual percepts, we need to define a formal relationship between sensory measurements and perceived motion. For this, we derive a Bayesian ideal observer that combines this sensory information with a prior distribution over 3-D motions. We will first consider only motion in the midsagittal plane ( $x_0 = 0$ ), such that uncertainty in  $x'$  and  $z'$  from the likelihood come out as independent (Equations 12 and 13).

The full likelihood function in real-world coordinates—that is, the conditional probability of the (transformed) velocity measurements  $\tilde{x}'$  and  $\tilde{z}'$  given the true velocities ( $x'$ ,  $z'$ )—is given by a 2-D Gaussian probability density function. For motion originating along the midsagittal plane, this is given by the product of two 1-D Gaussians,  $\mathcal{N}(\mu, \sigma^2)$ , where  $\mu$  and  $\sigma$  denote

the mean and standard deviation of a 1-D Gaussian, respectively. These likelihoods are given by

$$P(\ddot{x}'|x') = \mathcal{N}(x', \sigma_{x'}^2) \quad (14A)$$

$$P(\ddot{z}'|z') = \mathcal{N}(z', \sigma_{z'}^2), \quad (14B)$$

where  $\ddot{x}'$  and  $\ddot{z}'$  are derived from the measurements of retinal velocity in the left and right eyes (Equations 7–9). Note that we assume that  $h_L$  and  $h_R$  in these equations can be determined if one knows the interocular separation and has a reliable estimate of the object location from any combination of monocular and binocular cues (that is, the distance from each eye need not be derived from monocular distance information only).

We assume that the prior for slow speeds is isotropic in world velocities (a Gaussian with equal variance in all directions) and centered at 0 in  $x'$  and  $z'$ , as has been done previously (Lages, 2006, 2013; Wang et al., 2012; Welchman et al., 2008). We can then express the prior as

$$P(x') = \mathcal{N}(0, \sigma_p^2) \quad (15A)$$

$$P(z') = \mathcal{N}(0, \sigma_p^2), \quad (15B)$$

where  $\sigma_p^2$  is the variance of the prior.

The posterior distribution, according to Bayes's rule, results from multiplying the likelihood and the prior and renormalizing. For Gaussian likelihoods and priors, the posterior distribution also takes the form of a Gaussian, with mean and variance that can be computed according to standard formulas. The means of the posterior in  $x'$  and  $z'$  are given by

$$\hat{x}' = \alpha_{x'} \ddot{x}' \quad (16A)$$

$$\hat{z}' = \alpha_{z'} \ddot{z}', \quad (16B)$$

and the variances by

$$\sigma_{\hat{x}'}^2 = \alpha_{x'} \sigma_{x'}^2 \quad (17A)$$

$$\sigma_{\hat{z}'}^2 = \alpha_{z'} \sigma_{z'}^2, \quad (17B)$$

where  $\alpha_{x'}$  and  $\alpha_{z'}$  are “shrinkage factors” between 0 and 1 that govern how much the maximum-likelihood velocity estimates are shrunk toward zero (the mean of the prior). They are given here by

$$\alpha_{x'} = \frac{\sigma_p^2}{\sigma_{x'}^2 + \sigma_p^2} \quad (18A)$$

$$\alpha_{z'} = \frac{\sigma_p^2}{\sigma_{z'}^2 + \sigma_p^2}. \quad (18B)$$

The means in Equations 16A and 16B are denoted by  $\hat{x}'$  and  $\hat{z}'$  because they also correspond to the sensory estimate of each motion component determined by the posterior maximum, or maximum a posteriori (MAP) estimate. In brief, the estimated speeds correspond to the measured speeds in each direction, scaled toward zero by the shrinkage factor in each direction. Similarly, the posterior variance equals the variance of the sensory measurements also scaled by the shrinkage factor.

The full posterior distribution—that is, the probability of a given world velocity given a particular measured velocity—can therefore be written as

$$P(x'|x') = \mathcal{N}(x', \sigma_{x'}^2) \quad (19A)$$

$$P(z'|z') = \mathcal{N}(z', \sigma_{z'}^2). \quad (19B)$$

We can examine the trial-to-trial performance of the Bayesian ideal observer by deriving the sampling distribution of the MAP estimate—that is, the distribution of the estimates of a Bayesian ideal observer over multiple repeated presentations of a fixed stimulus. (The ideal observer exhibits variability because it receives a new set of noisy measurements on each trial). This distribution is given by

$$P(\hat{x}'|x') = \mathcal{N}(\alpha_{x'} x', \alpha_{x'}^2 \sigma_{x'}^2) \quad (20A)$$

$$P(\hat{z}'|z') = \mathcal{N}(\alpha_{z'} z', \alpha_{z'}^2 \sigma_{z'}^2). \quad (20B)$$

The variances of the ideal observer's estimates are scaled by  $\alpha_{x'}^2$  and  $\alpha_{z'}^2$  relative to the variance of a maximum-likelihood estimator (due to the fact that a Gaussian random variable scaled by  $\alpha$  will have its variance scaled by  $\alpha^2$ ). This shows that the ideal observer exhibits a reduction in variance even as it exhibits an increase in bias (in this case, bias toward slower speeds).

## Application of Bayes's rule to predict perceived motion at arbitrary x-z locations

For the case of motion occurring away from the midsagittal plane, we can derive the full covariance matrix of the ideal observer's estimates (which is not aligned with the  $x$ - and  $z$ -axes). We already have the noise covariance  $M$  of the sensory measurements from Equation 11. The covariance of the posterior of the Bayesian ideal observer (denoted by  $\Lambda$ , the covariance of  $\hat{x}'$  and  $\hat{z}'$ ) can be determined from this matrix and

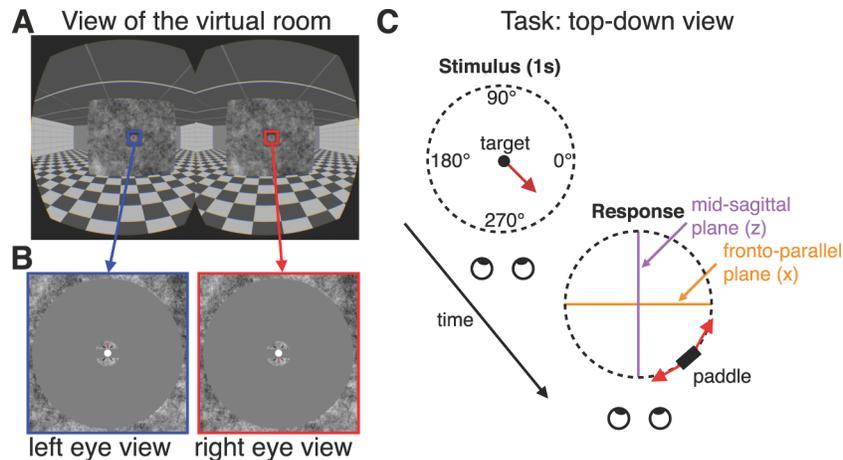


Figure 4 Stimulus and procedure for Experiment 1. (A) Participants wore a head-mounted display and viewed a stereoscopic virtual room with a planar surface in the middle. (B) Zoomed-in views of the left and right eyes' images show the critical aspects of the stimulus. Participants fixated nonius lines in the center of a circular aperture, and a virtual target (white sphere) appeared inside the nonius lines. (C) The target moved at a constant velocity in a random direction within the  $xz$ -plane (Stimulus). Afterwards, participants positioned a virtual paddle such that it would intersect the trajectory of the target (Response). The setting denoted by the black paddle in this example would result in a successful target interception.

the covariance of the prior (denoted as  $C$ , a diagonal matrix with variance in  $x'$  and  $z'$  of  $\sigma_p^2$ ):

$$\Lambda = (M^{-1} + C^{-1})^{-1}. \quad (21)$$

Given a pair of transformed velocity measurements  $\ddot{\mathbf{w}}' = (\ddot{x}', \ddot{z}')$ , the vector of posterior means in  $x'$  and  $z'$ —that is,  $\hat{\mathbf{w}}' = (\hat{x}', \hat{z}')$ , the MAP estimate in  $x'$  and  $z'$ —is then

$$\hat{\mathbf{w}}' = \Lambda M^{-1} \ddot{\mathbf{w}}'. \quad (22)$$

Here, the matrix  $S = \Lambda M^{-1}$  provides a joint shrinkage factor on the maximum-likelihood estimate analogous to the role played by  $\alpha$  in the previous section.

Lastly, the sampling distribution of the MAP estimate can be described as a 2-D Gaussian:

$$P(\hat{\mathbf{w}}' | \mathbf{w}') = \mathcal{N}(S\mathbf{w}', SMS^T), \quad (23)$$

where  $\mathbf{w}' = (x', z')$ . In the next sections, we describe the methods of a set of 3-D motion-perception experiments and compare the experimental results to the predictions of this ideal-observer model. Since the selection of axes over which to calculate motion is arbitrary, in all cases we convert world motion (whether the actual stimulus, the MAP estimate, or the actual participant response) to an angular-motion direction, with motion rightward defined as  $0^\circ$  and angles increasing counterclockwise (Figure 4C). In particular, we focus on systematic errors in the predicted and perceived direction of motion, and the impact of viewing distance, stimulus contrast, and lateral eccentricity on these errors.

## Experimental methods

In each experiment, observers viewed briefly presented stimuli moving with a random velocity in a random direction in the  $xz$ -plane, and indicated the perceived direction of motion. Raw data from all three experiments are provided in Supplementary Figures S1–S3.

### Experiment 1

The goal of Experiment 1 was to test the model predictions regarding the effects of viewing distance and stimulus contrast on perceptual errors (lateral bias and direction misreports) in the midsagittal plane using a naturalistic virtual-reality (VR) paradigm.

#### Participants

Seventy college-aged members of the University of Wisconsin–Madison community (43 women, 27 men) gave informed consent to complete the study, and 47 (26 women, 21 men) successfully completed all parts of the experiment. The participants that did not complete the study had difficulty either understanding the task or perceiving depth in the display ( $n = 19$ ) or wearing glasses inside the VR head-mounted display system ( $n = 4$ ). The experiment was carried out in accordance with the guidelines of the University of Wisconsin–Madison Institutional Review Board. Course credits were given in exchange for participation.

All participants had normal or corrected-to-normal vision and were screened for intact stereovision using the Randot Stereotest (Stereo Optical Company, Chicago, IL). To qualify for the study, participants were required to accurately identify all of the shapes in the Randot Form test, to identify the location of at least five out of 10 targets in the Randot Circle test, and to pass the suppression check. Although all participants passed the tests at these criteria, those with lower scores on the Form test (i.e., 5 or 6) were more likely to terminate their participation early (~50% of those who consented but did not complete the study).

### Apparatus

The experiment was controlled using MATLAB and the Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard, Pelli, Ingling, & Murray, 2007; Pelli, 1997) on a Macintosh computer and projected through an Oculus Rift DK1 (Oculus VR, Menlo Park, CA), which was calibrated using standard gamma calibration procedures. The Oculus Rift DK1 is a stereoscopic head-mounted VR system with an 18-cm LCD screen embedded in the headset providing an effective resolution of  $640 \times 800$  pixels per eye with a refresh rate of 60 Hz. The horizontal field of view is over  $90^\circ$  ( $110^\circ$  diagonal). Also embedded within the headset is a 1,000-Hz Adjacent Reality Tracker that relies upon a combination of gyros, accelerometers, and magnetometers to measure head rotation along the yaw, pitch, and roll axes with a latency of 2 ms. Note that translations of the head are not tracked by the device. Participants used a wireless keyboard to initiate trials and make responses.

### Stimulus and procedure

In a series of trials, participants were asked to indicate the perceived direction of motion of a target sphere that moved with a constant velocity in the virtual environment. The stimuli were presented in the center of a virtual room (346 cm in height, 346 cm in width, and 1,440 cm in depth). The virtual wall, ceiling, and floor were all mapped with different textures. These textures were included to facilitate better judgment of distances throughout the virtual space and the relative positions of the stimuli (Figure 4A).

The stimuli were otherwise similar to those used by Fulvio et al. (2015). In the center of the virtual room, there was a planar surface with a circular aperture ( $7.5^\circ$  in radius). The planar surface was mapped with a  $1/f$  noise pattern that was identical in both eyes to aid vergence. In addition, nonius lines were embedded within a small  $1/f$  noise patch near the center of the aperture. All stimulus elements were antialiased to achieve subpixel resolution. The background seen through the aperture was midgray (Figure 4B).

The planar surface was positioned in the room at one of two viewing distances from the observer's location: 90 cm ( $n = 15$  participants) or 45 cm ( $n = 32$  participants). Participants were instructed to fixate the center of the aperture. However, they were free to make head movements, and when they did so, the display updated according to the viewpoint specified by the yaw, pitch, and roll of their head. Translations of the head did not affect the display, such that stimulus viewing distance remained constant.

On each trial, a white sphere (target) of  $0.25^\circ$  in diameter appeared at the center of the aperture and then followed a trajectory defined by independently chosen random speeds in the  $x$  (lateral) and  $z$  (motion-in-depth) directions, with no change in  $y$  (vertical direction), before disappearing. The motion trajectory always lasted for 1 s. Velocities in  $x$  and  $z$  were independently chosen from a 2-D Gaussian distribution ( $M = 0$  cm/s,  $SD = 2$  cm/s) with imposed cutoffs at 6.1 and  $-6.1$  cm/s. This method resulted in motion trajectories whose directions spanned the full  $360^\circ$  space (Figure 4C, left side). Thus, the target came toward the participant (approaching) and moved back behind fixation away from the participant (receding) on approximately 50% of trials each. It is important to note that since  $x$  and  $z$  motion were chosen randomly and independently, the amount of perceived lateral movement on each trial did not carry information about the amount of motion in depth, and vice versa. The target was rendered under perspective projection, so that both monocular (looming) and binocular cues to motion in depth were present.

Participants indicated the perceived target trajectory using a “3-D Pong” response paradigm (Fulvio et al., 2015). After the target disappeared, a 3-D rectangular block (paddle), whose faces also consisted of a  $1/f$  noise pattern, appeared at the edge of the aperture. The paddle dimensions were  $0.25 \text{ cm} \times 0.5 \text{ cm} \times 0.25 \text{ cm}$ . Participants were asked to extrapolate the target's trajectory and adjust the paddle's position such that the paddle would have intercepted the target if the target had continued along its trajectory. The paddle's position could be adjusted along a circular path that orbited the fixation point in the  $xz$ -plane using the left and right arrow keys of the keyboard (Figure 4C, right side). As the participant moved the paddle through the visual scene, the paddle was rendered according to the rules of perspective projection. Thus, the stimuli were presented and the responses were made in the same 3-D space. By asking participants to extrapolate the trajectory, we prevented them from setting the paddle to a screen location that simply covered the last seen target location. We did not ask participants to retain fixation during the paddle-adjustment phase of the trial. When participants were satisfied with the paddle setting, they resumed fixation and pressed the space bar to initiate

a new trial. Supplementary Movie S1 demonstrates the general procedure.

The target had variable contrast, presented at one of three Weber values—100% (high), 15% (mid), and 7.5% (low)—which were counterbalanced and presented in pseudorandom order.

Participants carried out 10 to 15 practice trials in the presence of the experimenter to become familiar with the task. All participants completed the experimental trials in one session that was self-paced. The data reported here were collected as part of a larger study. Within the study, these data comprise one block of trials, which took on average 5–12 min for each participant to complete and contained an average of 225 trials (an entire session was 30–45 min with breaks throughout). No feedback was provided for either the practice or experimental trials.

### Data analysis

To examine biases in the perceived direction of motion, we computed the mean angular error for each participant and each unique stimulus condition (viewing distance and target contrast). Errors were calculated as the angular distance of the reported direction relative to the stimulus direction in the  $xz$ -plane. We analyzed the data to determine whether this angular error tended to be toward the fronto-parallel plane (lateral bias) or the midsagittal plane (medial bias; see Figure 4C). We assigned positive values to medial errors and negative values to lateral errors such that the average would indicate the overall directional bias.

To examine the frequency of motion-direction errors, we computed the percentage of trials on which paddle settings were made on the opposite side of either the fronto-parallel or the midsagittal plane. Responses on the opposite side of the fronto-parallel plane (approaching vs. receding) were considered depth-direction confusions. Responses made on the opposite side of the midsagittal plane (leftward vs. rightward) were considered lateral-direction confusions.

Statistical effects were tested through an analysis of variance (ANOVA) evaluated on generalized linear model fits to the individual trial data for mean lateral bias and to the individual subject proportions for direction-confusion analyses. The model incorporated viewing distance as a between-subjects fixed effect and target contrast as a within-subject (repeated-measures) fixed effect. The model intercepts were included as random subject effects. Follow-up tests consisted of Bonferroni-corrected  $t$  tests for multiple comparisons.

## Experiment 2

To examine whether the motion-direction confusions measured in Experiment 1 were particular to the VR

setup, we compared these results to those of a second experiment conducted on a traditional stereoscopic display.

### Participants

Three adults participated in the experiment. All had normal or corrected-to-normal vision. One participant (a man, age 23 years) was unaware of the purpose of the experiment and had limited psychophysical experience. The remaining two participants (authors JWP and BR; men, age 34–35 years) had extensive psychophysical experience. The experiment was undertaken with the written consent of each observer, and all procedures were approved by the University of Texas at Austin Institutional Review Board.

### Apparatus

The experiment was performed using a similar setup to Experiment 1; however, in this case the stimuli were presented on two 35.0 cm  $\times$  26.3 cm CRT displays (ViewSonic G90fB, one for each eye; 75 Hz; 1,280  $\times$  1,024 pixels) at a single viewing distance of 90 cm (21.2°  $\times$  16.3° of visual angle). Left- and right-eye half images were combined using a mirror stereoscope. The luminance of the two displays was linearized using standard gamma-correction procedures, and the mean luminance was 50.6 cd/m<sup>2</sup>.

### Stimulus and procedure

As in Experiment 1, all stimuli were presented within a circular midgray aperture (1° radius) that was surrounded by a  $1/f$  noise texture at the depth of the fixation plane (90 cm) to help maintain vergence. No virtual room was present. Additionally, a small square fixation point was placed at the center of the display. The fixation point was surrounded by horizontal and vertical nonius lines, and was placed on a circular 0.1° radius  $1/f$  noise pattern.

Rather than a single target, a field of randomly positioned dots moving in the  $xz$ -plane was presented on each trial. The positions of the dots were constrained to a single plane fronto-parallel to the display (i.e., perpendicular to the observer's viewing direction). The initial disparity of the plane was consistent with a distance of 93 cm (3 cm behind the fixation plane). The plane then moved for 500 ms with  $x$  and  $z$  velocities independently and uniformly sampled from an interval of  $-4$  to  $4$  cm/s, corresponding to a maximum possible binocular disparity of 0.21° (uncrossed) relative to the fixation plane.

Each moving dot had a 200-ms limited lifetime to prevent tracking of individual stimulus elements. Dot radius was 0.11 cm and dot density was  $\sim 74$  dots/deg<sup>2</sup>. Both dot size and dot density changed with distance

to the observer according to the laws of perspective projection. Dots had variable contrast, presented at one of three Weber values (7.5%, 15%, or 60%). Weber contrast was computed as the luminance difference between the dots and the background expressed as a percentage of the background luminance level. Half of the dots were darker, and the other half brighter, than the midgray background.

The stereoscope was initially adjusted so that the vergence demand was appropriate for the viewing distance, given a typical interocular distance. Prior to each session, each participant made further minor adjustments so that the nonius lines at fixation were aligned both horizontally and vertically, and vergence was comfortable. Participants were instructed to maintain fixation for the duration of each experimental trial.

Trials proceeded as described for Experiment 1, except that participant responses were made differently. After the dots disappeared, a circle and a line were presented on-screen, where one of the line endpoints was fixed to the center of the circle and the participant could adjust the other line endpoint with a computer mouse. Participants were instructed to treat this as a top-down view of the stimulus (see Figure 4C), and to adjust the line such that the angle was consistent with the trajectory of the dots. We verified in pilot experiments that this method produced consistent, reproducible estimates. As in Experiment 1, no feedback concerning performance was provided.

### Data analysis

Data were analyzed in the same manner as in Experiment 1.

## Experiment 3

To test model predictions for stimuli presented at locations at eccentricity—away from the midsagittal plane—we conducted a third experiment using the same VR display as described in Experiment 1.

### Participants

Twenty-two college-aged members of the University of Wisconsin–Madison community gave informed consent to complete the study. One participant did not complete the study because of difficulty in perceiving depth in the display, despite passing the stereovision screening. The remaining 21 participants completed all aspects of the experiment. The experiment was carried out in accordance with the guidelines of the University of Wisconsin–Madison Institutional Review Board. Course credits were given in exchange for participation. All participants had normal or corrected-to-normal

vision and were screened for intact stereovision using the Randot Stereotest in order to meet the criteria outlined for Experiment 1.

### Apparatus

The apparatus was the same as that described for Experiment 1.

### Stimulus and procedure

The stimulus and procedure were similar to those of Experiment 1, with the exception that the planar surface in the center of the room had three circular apertures rather than just one. As in Experiment 1, one of the apertures appeared at the center of the planar surface directly in front of the participants ( $7.5^\circ$  radius). The other two apertures were located  $20^\circ$  to the left and right of the central location. These two apertures were slightly larger ( $10.5^\circ$  radius) in order to ensure adequate visibility of the stimulus. All three apertures appeared on every trial, and the background seen through the aperture was black, which increased the contrast of the stimuli and further improved visibility in the periphery (Figure 4B).

The planar surface was positioned in the virtual room at 45 cm from the participants. Participants were instructed to fixate the center of the central aperture on every trial, even when a target appeared at one of the peripheral locations. This instruction served to minimize head rotation, not eye movements per se. Recall that model predictions critically depend on stimulus location in 3-D space (relative to the observer), not stimulus position on the retina.

On each trial, a white sphere appeared at the center of one of the three apertures randomly and counterbalanced across trials. To ensure that the peripheral targets were clearly visible while participants fixated the central aperture, these targets were rendered with a diameter of  $0.5^\circ$  versus the  $0.25^\circ$  in the central location. Targets were always presented at full contrast corresponding to a  $\sim 65,000\%$  Weber value—that is, white ( $92.4 \text{ cd/m}^2$ ) on a black ( $0.14 \text{ cd/m}^2$ ) background. All other aspects of the target's motion were identical to those in Experiment 1.

Participants indicated the perceived target trajectory using a “3-D Pong” response paradigm as in Experiment 1 at each of the three aperture locations. They were free to move their eyes to the three apertures during the response phase of the trial. Participants carried out 10 to 15 practice trials in the presence of the experimenter to become familiar with the task. All participants completed the experimental trials in one session. No feedback was provided in either the practice or experimental trials. All participants completed 360 experimental trials, divided into three blocks to allow for breaks from wearing the head-mounted display.

Across participants, the average time to complete each of the three blocks was 5–8 min, for a total of 15–24 min of active experimental time in a session.

### Data analysis

Data were analyzed in the same manner as in Experiment 1.

### Goodness of fit

To fit our model to the experimental data, we assumed that the participant response for motion direction on each trial was based on the polar angle of the MAP estimate, such that

$$\hat{\theta} = \arctan\left(\frac{\hat{x}'}{\hat{z}'}\right). \quad (24)$$

Thus, we fit the distribution of responses over  $\hat{\theta}$  given the stimulus velocities, which is given by the offset normal distribution (Jammalamadaka & Sengupta, 2001). This distribution is written as

$$P(\hat{\theta}|x', z') = \frac{\phi(l)}{j\gamma_{x'}\gamma_{z'}\sqrt{2\pi}} \left[ k \frac{\Phi(k)}{\phi(k)} + 1 \right], \quad (25)$$

where

$$j = \frac{\cos^2 \hat{\theta}}{\gamma_{x'}^2} + \frac{\sin^2 \hat{\theta}}{\gamma_{z'}^2} \quad (26)$$

$$k = \left( \frac{\hat{x}' \cos \hat{\theta}}{\gamma_{x'}^2} + \frac{\hat{z}' \sin \hat{\theta}}{\gamma_{z'}^2} \right) j^{-1/2} \quad (27)$$

$$l = \left( \frac{\hat{x}'^2}{\gamma_{x'}^2} + \frac{\hat{z}'^2}{\gamma_{z'}^2} \right)^{1/2} \quad (28)$$

and  $\phi(\cdot)$  and  $\Phi(\cdot)$  denote, respectively, the standard normal probability density function and the cumulative density function. Here,  $\gamma_{x'}^2 = \alpha_{x'}^2 \sigma_{x'}^2$  and  $\gamma_{z'}^2 = \alpha_{z'}^2 \sigma_{z'}^2$  are the variances of the sampling distribution of the MAP estimate for  $x'$  and  $z'$ , respectively (Equations 20A, B).

We found the maximum-likelihood estimate of the model parameters by numerically optimizing the log likelihood over all trials  $i$ :

$$\log P(\{\hat{\theta}_i\} | \{x'_i, z'_i\}) = \sum_i \log P(\hat{\theta}_i | x'_i, z'_i) \quad (29)$$

for the standard deviations of the measurement noise ( $\sigma_{n_p}$ , in deg/s) and the prior ( $\sigma_p$ , in cm/s and assumed to be isotropic). Using this method, we fit the model

to the motion-direction responses of each individual participant in each experiment. Note that the parameter for the prior was fitted for each participant based on the assumption that individuals do not have identical priors. For Experiments 1 and 2, the measurement noise for each stimulus contrast was fit independently.

Along with the noise parameters, the goodness of fit  $g$  of this model was assessed by comparing the log likelihood of the best-fit parameters to the log likelihood achieved if the prior was assumed to be uniform (not Gaussian). This comparison was calculated in terms of the bits per trial gained by fitting with a zero-mean Gaussian prior:

$$g = \left( \frac{\log P(\{\hat{\theta}_i\} | \{x'_i, z'_i\}) - \log P_u(\{\hat{\theta}_i\} | \{x'_i, z'_i\})}{I} \right) \frac{1}{\log e}, \quad (30)$$

where  $P_u(\cdot)$  denotes the likelihood associated with the uniform prior and  $I$  denotes the total number of trials for a given participant.

To compare the model-predicted errors in motion perception to the experimental data, we used various descriptive statistics as defined in the experimental methods: mean angular error (lateral/medial bias), motion-in-depth direction confusions, and lateral-direction confusions. For the model predictions, we computed the predicted errors using each participant's fitted parameters and assumed a typical interocular separation of 6.4 cm, a uniform sampling of stimulus directions, and an average stimulus speed consistent with each experiment.

## Results

### Comparing model predictions to observed behavior

The average and standard deviation of the fitted noise parameters for each experiment are shown in Table 1. Note that for Experiments 1 and 2, separate noise parameters were determined for each stimulus contrast level. As expected, in these experiments the best-fit noise increased monotonically (but not linearly) with decreasing contrast. Quantitative predictions for how velocity noise should vary with contrast have been derived previously (Hürlimann et al., 2002), and in particular have suggested that the internal representation of stimulus contrast may be highly nonlinear, particularly when a broad range of contrasts are considered.

A goodness-of-fit measure  $g$  is also provided for each experiment in Table 1. Across all three experiments, between 0.29 and 0.87 bits/trial was gained on average through the inclusion of a Gaussian slow-motion prior. For Experiment 3, six (out of 21) participants were best

Experiment 1	100% $\sigma_{n_{\beta'}}$	15% $\sigma_{n_{\beta'}}$	7.5% $\sigma_{n_{\beta'}}$	$\sigma_p$	$g$
$M$ (SD) at 45 cm	0.66 (0.45)	0.88 (0.50)	1.18 (0.76)	7.75 (3.86)	0.29 (0.30)
$M$ (SD) at 90 cm	0.52 (0.30)	0.66 (0.31)	0.96 (0.38)	10.82 (4.16)	0.81 (0.47)
Experiment 2	60% $\sigma_{n_{\beta'}}$	15% $\sigma_{n_{\beta'}}$	7.5% $\sigma_{n_{\beta'}}$		
$M$ (SD)	0.27 (0.35)	0.31 (0.35)	0.41 (0.45)	5.05 (6.42)	0.87 (0.52)
Experiment 3	$\sigma_{n_{\beta'}}$				
$M$ (SD)	1.14 <sup>a</sup> (1.18) <sup>a</sup>	—	—	10.13 <sup>a</sup> (7.70) <sup>a</sup>	0.31 (0.81)

Table 1. Noise estimates from fitting the Bayesian ideal-observer model to each participant's responses. For each experiment, mean ( $M$ ) and standard deviation ( $SD$ , both in deg/s) of the noise estimates across participants for each stimulus contrast, and the prior ( $\sigma_p$ , in cm/s), are shown. Estimates were computed separately for the participant groups from the two viewing distances in Experiment 1 (45 and 90 cm). The goodness of fit  $g$  for each complete dataset is reported, as described in Equation 30. <sup>a</sup>Parameters from six outlier participants were not included in the calculation of the mean and standard deviation.

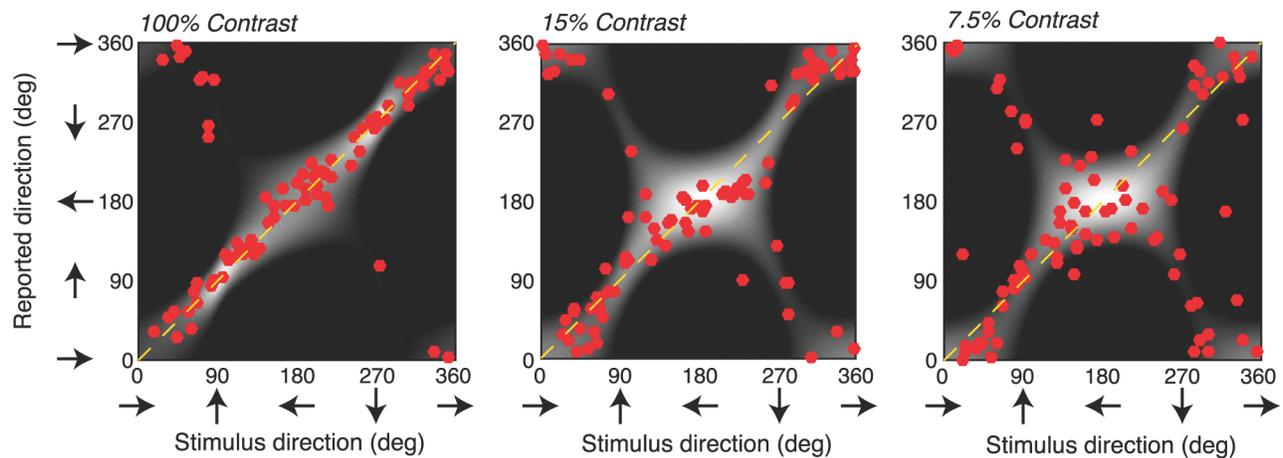


Figure 5 Model predictions and human performance for an example participant in Experiment 1. Red data points reflect individual trials with a randomly chosen trajectory on each trial, but a shared starting location 45 cm directly ahead of the participant. The grayscale color map indicates the probability density of the sampling distribution of the maximum a posteriori estimate for each stimulus direction, and each panel is normalized so that the probability densities span the full color map. For this participant, the standard deviation of the prior was estimated as 3.76 cm/s. Standard deviations of the retinal measurement noise were estimated as 0.42°/s, 1.08°/s, and 1.27°/s for stimuli at 100%, 15%, and 7.5% contrast, respectively.

described as having a flat prior (that is, essentially 0 bits/trial were gained with the Gaussian prior, and the best-fit  $\sigma_p > 1,000$  cm/s). The noise-parameter mean and standard deviations in the table exclude the fits to these participants (but all participants are included in the goodness of fit and all subsequent analyses). Even excluding these six participants, the estimated sensory noise for the participants in Experiment 3 was greater than for the high-contrast condition in the other experiments. Recall that Experiment 3 had three potential stimulus eccentricities: Although only the central position was used to fit the model, it seems reasonable that the demand to attend to all three locations may have increased the sensory uncertainty in this experiment.

We can compare model predictions with behavioral performance for individual participants by plotting the predicted MAP sampling distributions along with the trial-by-trial data. Figure 5 shows this comparison for an example participant from Experiment 1. We plot presented stimulus direction on the horizontal axes and predicted/reported direction on the vertical axis. Arrows along the axes indicate these motion directions from a top-down view (e.g., a downward arrow corresponds to approaching motion). By convention, rightward motion is defined as 0° and increases counterclockwise (compare to Figure 4C). The figure axes depict circular data, such that 0° and 360° both represent rightward motion. The red circles reflect the behavioral performance, depicting reported motion direction

(vertical axis) as a function of presented motion direction (horizontal axis). The grayscale data reflect the model predictions for this observer, showing the probability density of the MAP sampling distribution for each motion direction.

A few features stand out in these data. First, if reported motion direction always matched presented direction, data points should cluster along the positive diagonal (yellow dashed line). In inspecting performance at high contrast (left panel) it is clear that a considerable proportion of trials reflect this relationship for both model predictions and human behavior. As contrast is reduced (middle and right panels), performance becomes less accurate. In particular, both the predicted and reported motion directions start to cluster around  $180^\circ$  and  $0^\circ/360^\circ$ , representing an increase in the prediction/reporting of lateral motion directions at the expense of motion-in-depth directions ( $90^\circ/270^\circ$ ). Second, those predictions and data points that fall away from the positive diagonal seem to cluster around the negative diagonal. What do such points indicate? These data points correspond to trials in which the observer accurately judged the lateral component of the 3-D motion trajectory but incorrectly judged the motion-in-depth component, reporting approaching trajectories as receding and vice versa. Third, model predictions and human performance seem to be in agreement, at least qualitatively. In particular, as stimulus contrast is reduced, precision decreases and off-diagonal points increase. In the next sections, we will explore these errors further and assess quantitative agreement between the model predictions and the results of three perceptual experiments.

### Predicted and observed biases toward lateral motion in the midsagittal plane vary with stimulus distance and contrast

Recall that previous perceptual experiments have demonstrated that observers tend to overestimate the angle of approach of objects (the lateral bias). That is, an object on a trajectory toward the observer tends to be perceived as moving more laterally than the true stimulus trajectory. Figure 6 summarizes the average angular lateral bias (predicted and observed) across all observers from Experiment 1 for each viewing condition. Bars indicate the average signed error between the stimulus and either the percept predicted by the model (Figure 6A) or the measured participant responses from the experiment (Figure 6B). Larger values of this error indicate larger lateral biases (see Experimental methods). The overall effects of stimulus distance and contrast are well matched to the model predictions. The model (Figure 6A) predicts a decrease in the lateral bias with decreased viewing distances (i.e., 90 cm vs. 45 cm). It also predicts an increase in

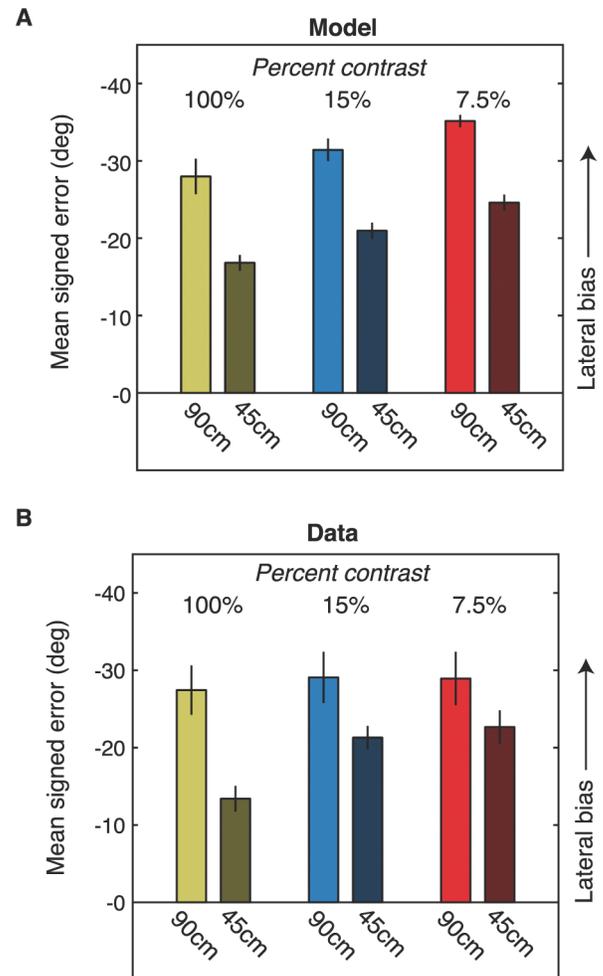


Figure 6 Comparison between model predictions and human lateral bias in Experiment 1. (A) Mean signed error in predicted perceived target direction for viewing three target contrast levels at two viewing distances. Negative values (increasing on the ordinate) correspond to reports that are laterally biased. (B) Results for the 47 participants ( $n = 15$  for 90 cm and  $n = 32$  for 45 cm) who took part in Experiment 1 (viewing the stimulus within the virtual-reality environment), plotted in the same format as (A). Error bars correspond to  $\pm 1$  standard error of the mean.

the lateral bias for lower contrast stimuli. Both of these effects are reflected in the observed errors in the behavioral experiment (Figure 6B), although an effect of contrast is only apparent at the closer viewing distance.

A two-way ANOVA performed on the experimental data showed a significant interaction between viewing distance and target contrast,  $F(2, 10494) = 5.8$ ,  $p < 0.01$ . Multiple comparisons revealed a significant increase in perceptual bias at the greater viewing distance compared to the smaller viewing distance for the mid and high target contrast levels ( $p < 0.01$ ). The difference in perceptual bias at the two viewing

distances for the low target contrast was not significant ( $p > 0.05$ ).

While a previous model did predict an effect of viewing distance (Welchman et al., 2008), prior experimental studies have concluded that distance does not modify the lateral bias (Harris & Dean, 2003; Poljac, Neggers, & van den Berg, 2006). Until now, this inconsistency between model and data did not have a clear explanation. However, as demonstrated in Figure 6B, the magnitude of the difference between viewing distances interacts with other properties of the stimulus uncertainty (here shown as contrast, but generally summarized as the variance of the likelihood). Thus, it is possible that some experimental setups would reveal a distance effect and others might not, particularly with relatively small sample sizes such as those used in the previous studies (three and six participants, respectively). While the model predicts a viewing-distance effect at all contrasts in the current study, the magnitude of the effect does decrease slightly with decreasing contrast (from  $11.1^\circ$  at 100% contrast to  $10.6^\circ$  at 7.5% contrast).

In addition to the dependence of the lateral bias on viewing distance, the model predicts a dependence on stimulus eccentricity. It predicts that the relative stimulus uncertainty in depth (and therefore the lateral bias) should be reduced when an object is located off to the left or right rather than directly in front of the observer (Figures 1 and 3). We will return to this prediction in a later section.

### Misperceptions in motion direction in the midsagittal plane

Recent work has shown that motion-trajectory judgments tend to be subject to direction confusions for approaching and receding motion (Fulvio et al., 2015). That is, observers sometimes report that approaching stimuli appear to move away, and vice versa. Such dramatic errors are surprising. Within the same paradigm, observers rarely if ever report that leftward-moving stimuli appear to move rightward, and vice versa. Can motion-in-depth reversals be explained by our Bayesian model? We first examined this question by plotting the full sampling distribution of the MAP for two example stimuli: motion directly toward an observer and motion directly to the right (the left panels of Figure 7A and 7B, respectively, with model parameters from the participant plotted in Figure 5). Specifically, for each example stimulus we show a heat map of this distribution, with  $x'$  plotted on the horizontal axis and  $z'$  plotted on the vertical axis. These plots demonstrate that a large percentage of the sampling distribution for a stimulus moving toward an observer can occur for trajectories that recede in depth. In other words, the variance of the MAP sampling

distribution in the  $z$  direction can be large enough that it extends into the opposite direction of motion. For rightward motion, however, very little of the distribution occurs for leftward trajectories. To further examine the percentage of trials in which observers are predicted to misreport motion direction, we converted the trajectories in the sampling distribution of the MAP to direction angles and replotted the normalized frequency in polar coordinates as a function of motion direction (Figure 7A and 7B, right panels). Nonzero values in the opposite direction of motion (away or leftward) indicate that the model predicts that a certain percentage of trials will include direction confusions.

Next, we examined the effects of distance and contrast on predicted and observed direction confusions, averaging across all directions of motion in the world. The Bayesian model predicts that direction confusions for motion in depth (Figure 7C) will greatly exceed lateral-motion confusions (Figure 7D). Each bar represents the predicted percentage of trials in which direction will be confused, and the dashed line indicates chance performance (50%).

For motion-in-depth confusions, the model predicts that direction confusions will decrease with reduced viewing distance (90 cm vs. 45 cm). The model also predicts that direction confusions will increase as sensory uncertainty increases (contrast decreases from 100% to 15% to 7.5%), most markedly at the smallest viewing distance (dark bars). The upper right-hand panel in Figure 7C shows the results from Experiment 1, plotted in the same manner as the model predictions. The overall effects of stimulus distance and contrast are well matched to the model predictions.

A two-way ANOVA conducted on the data from Experiment 1 revealed a main effect of viewing distance on human performance,  $F(1, 135) = 7.8$ ,  $p < 0.01$ , as well as a main effect of target contrast,  $F(1, 135) = 26.79$ ,  $p < 0.01$ , with a reduction in direction confusions for object motion nearer to the head and for higher target contrasts. There was also a significant interaction between viewing distance and target contrast,  $F(2, 135) = 4.4$ ,  $p = 0.014$ . Multiple comparisons revealed that direction confusions significantly increased for all target contrast levels ( $p < 0.01$  low and high;  $p = 0.013$  mid) as the viewing distance doubled from 45 cm to 90 cm.

Because direction confusions might seem surprising, we compared these results to a second experiment (Experiment 2, lower right-hand panel of Figure 7C). This experiment used a standard stereoscopic display and a random-dot stimulus. Note that Experiment 2 included a contrast manipulation, but stimuli were always presented at one distance, and the high-contrast condition was 60% rather than 100% contrast. As predicted, a one-way ANOVA on the data from Experiment 2 revealed a main effect of target contrast,  $F(2, 4) = 160.99$ ,  $p < 0.01$ .

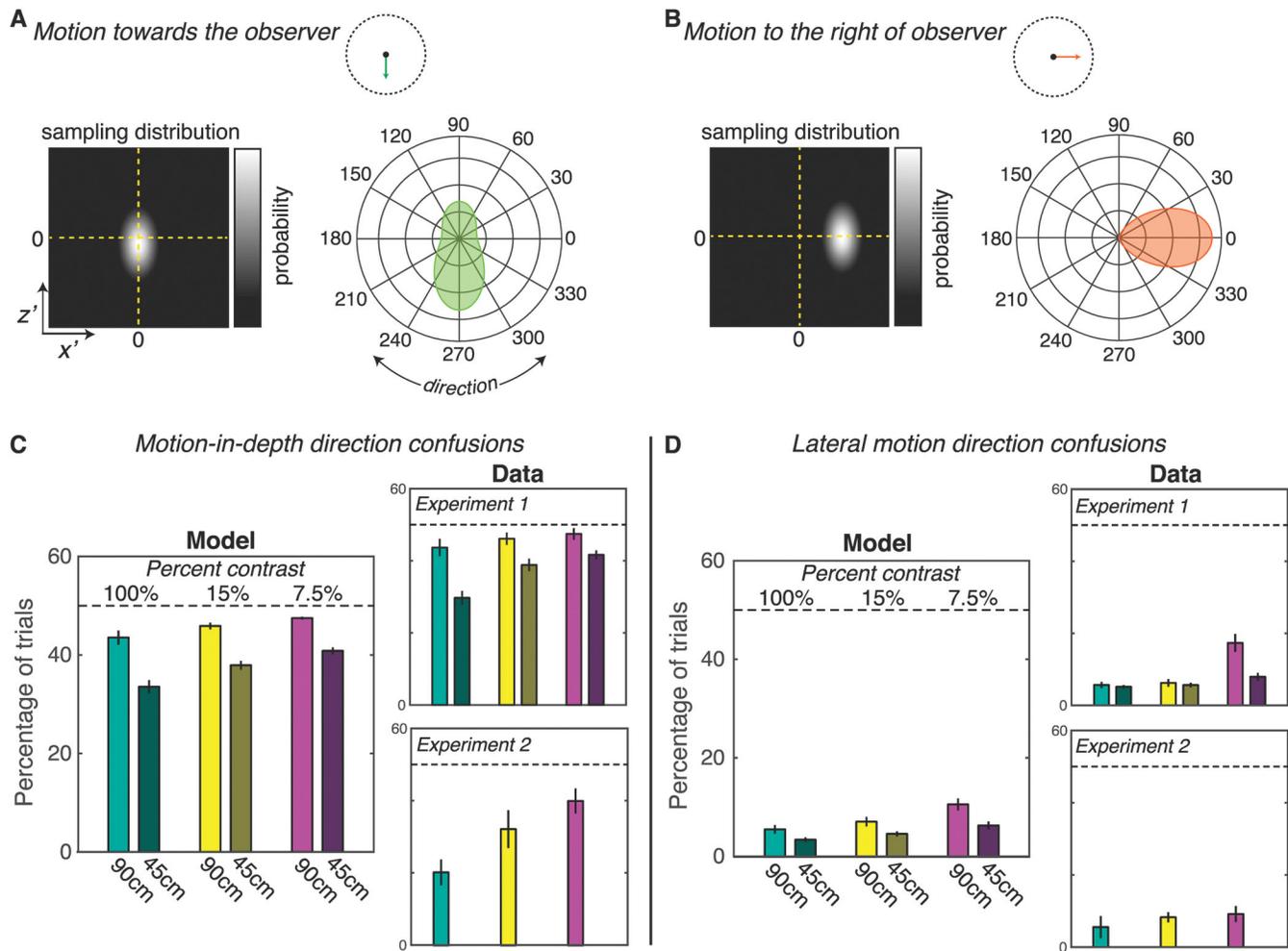


Figure 7 Direction confusions for motion in depth and lateral motion. (A, B) Illustrations of the predicted sampling distribution of the maximum a posteriori estimates for motion directly toward an observer (A) and directly to the right of an observer (B) in Cartesian and polar coordinates. Model parameters used were from the same example participant shown in Figure 5 (45-cm viewing distance, 7.5% contrast). (C, D) Predictions of the model and experimental results for motion-in-depth confusions (C) and lateral-motion confusions (D). Experiments 1 and 2 are shown in separate panels. Error bars correspond to  $\pm 1$  standard error of the mean.

The model predicts that lateral motion-direction confusions will be much less frequent but will be similarly affected by viewing distance and stimulus contrast (Figure 7D, left panel). That is, in the fronto-parallel plane, direction confusions will decrease with reductions in viewing distance and increase with reductions in stimulus contrast. These predicted effects were present in both experiments (Figure 7D, right panels). A two-way ANOVA on the data from Experiment 1 revealed a main effect of viewing distance,  $F(1, 135) = 17.35$ ,  $p < 0.01$ . The interaction between viewing distance and contrast was also statistically significant,  $F(2, 135) = 9.52$ ,  $p < 0.01$ . Follow-up comparisons revealed that direction confusions significantly increased with viewing distance for the lowest contrast stimulus ( $p < 0.01$ ).

Although the average percentage of lateral misperceptions was highest in the low-contrast

condition of Experiment 2, the effect of stimulus contrast was not statistically significant,  $F(2, 6) = 3.4$ ,  $p = 0.05$ .

To summarize, while overt confusions in the direction of motion seem surprising on their own, they are clearly predicted by the same Bayesian motion-perception model that accounts for other perceptual phenomena.

### 3-D motion perception outside of the midsagittal plane

The previous sections have considered motion trajectories originating in the midsagittal plane. Of course, in the real world, stimuli need not be confined to this plane and may originate in any location relative to the observer. While uncertainty in  $z'$  is typically much larger than uncertainty in  $x'$  for the same location in

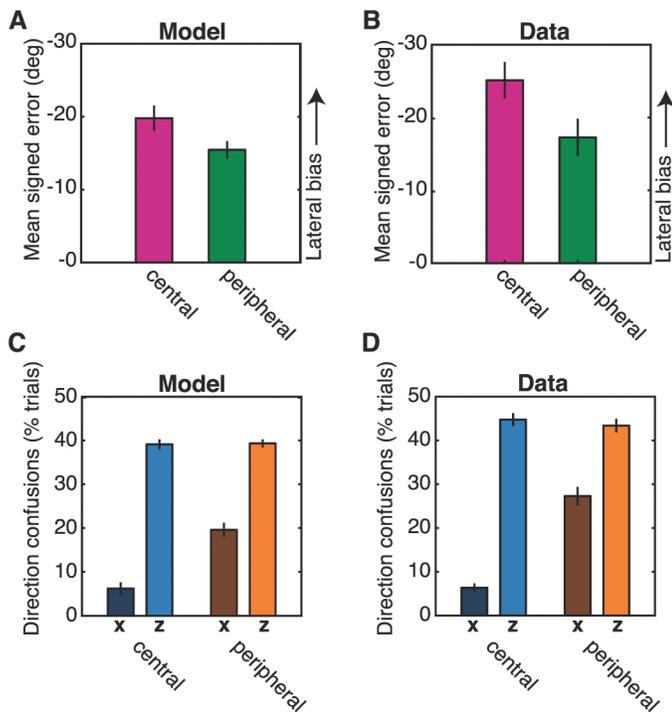


Figure 8 Lateral bias and direction confusions in central and peripheral locations. (A, B) Lateral-bias predictions of the model (A) and experimental results (B) for stimuli present in the midsagittal plane (central) and 20° to the left or right (peripheral). (C, D) Predictions of the model (C) and experimental results (D) for lateral-motion confusions (x) and motion-in-depth confusions (z). Error bars correspond to  $\pm 1$  standard error of the mean.

the midsagittal plane, the relative uncertainty decreases away from that plane (Figure 3). In fact, at an angle of 45° away from that plane the relative uncertainty becomes the same, predicting unbiased estimates of motion trajectory. Beyond 45° the relationship reverses, such that the model will predict a medial rather than a lateral bias. Another way to think about this is that the axis of maximal uncertainty shifts from being aligned with the z-axis in the midsagittal plane to become aligned with the x-axis for motion originating directly to the left or right of the observer (see Figure 1). Because of this, estimated motion trajectories predicted by the model will differ between midsagittal and peripheral motion.

In Experiment 3, we tested whether the observed lateral bias and motion-direction confusions are affected by stimulus eccentricity, in accordance with model predictions. Figure 8A and 8C shows the model predictions for lateral bias and direction confusions for motion trajectories originating in the midsagittal plane (central) and 20° to the left or right (peripheral). At eccentricity, the lateral bias is predicted to decrease (Figure 8A); the percentage of motion-direction confusions is predicted to increase for lateral motion

(x) but stay largely the same for motion in depth (z; Figure 8C). These model predictions are qualitatively similar to the experimental data (Figure 8B and 8D). Note that for this experiment, the model parameters were fitted to the central data for each participant, and then peripheral predictions were generated based on these parameters.

A paired-sample *t* test on the experimental data revealed significantly less lateral bias in response to peripheral compared to central targets,  $t(20) = -2.5$ ,  $p = 0.02$ , with a difference of 7.9°. There was a small decrease in motion-in-depth direction confusion at the peripheral locations of  $\sim 1.38\%$  on average, but this difference was not significant,  $t(20) = 0.78$ ,  $p > 0.05$ . By contrast, there was a substantial and significant increase in lateral motion-direction confusion (20.9% on average) at the peripheral locations,  $t(20) = -10.82$ ,  $p < 0.01$ , as predicted by the model.

### Additional biases

In addition to the lateral bias and depth-direction confusions, other researchers have documented that motion trajectories towards the observer have some amount of “privileged” perceptual processing (Lin, Franconeri, & Enns, 2008; Lin, Murray, & Boynton, 2009; Schiff, Caviness, & Gibson, 1962). Indeed, our prior work has shown a bias to report motion-in-depth stimuli as approaching rather than receding, or vice versa, depending on the specific appearance of the stimulus (Cooper, van Ginkel, & Rokers, 2016; Fulvio et al., 2015). These biases may be related to prior work which suggests that observers perceive lower contrast stimuli as farther away than high-contrast stimuli (Doshier, Sperling, & Wurst, 1986; Farnè, 1977; O’Shea, Blackburn, & Ono, 1994; Schwartz & Sperling, 1983). However, given that the contrast of the stimuli in the motion-in-depth experiments did not vary within a given trial, the exact nature of this relationship remains to be explored. Figure 9 illustrates how these approaching/receding biases manifest in the current set of experiments. Each panel shows the probability density of response directions, averaged over all participants from Experiment 1, for stimuli that moved toward four different quadrants: rightward (red), leftward (yellow), approaching (green), and receding (blue). Figure 9A shows the responses for the high-contrast (100%) smaller viewing distance (45 cm): The reported motion directions tended to generally fall within the stimulus quadrant, but some motion-in-depth direction confusions are clearly evident. Figure 9B and 9C shows the responses for two conditions in which participants, on average, were biased to report that motion was receding (B) and approaching (C). The current Bayesian model does not predict approaching or receding biases because

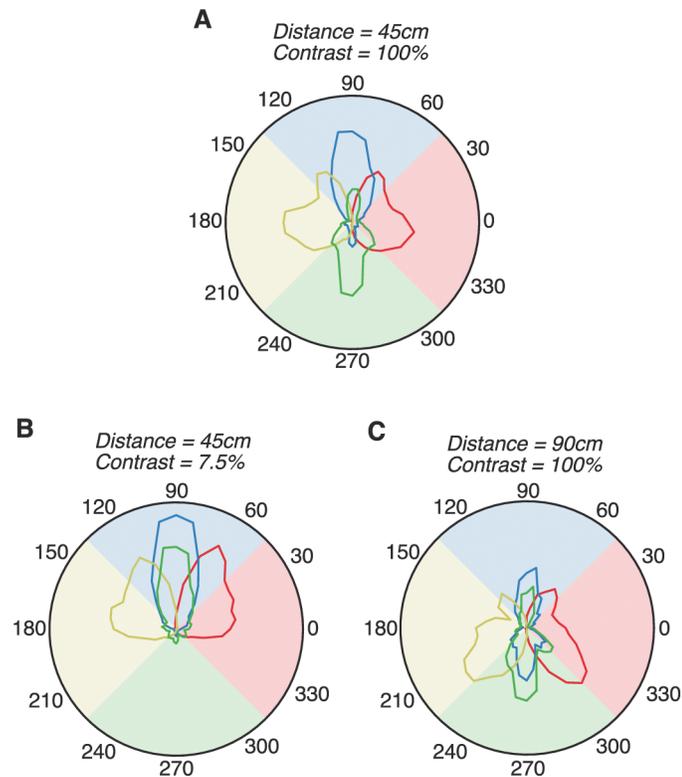


Figure 9 Approaching and receding biases under different viewing conditions. Each panel illustrates the probability density of responses for all stimulus motions falling within the indicated quadrants: rightward (315°–45°), receding (45°–135°), leftward (135°–225°), and approaching (225°–315°). Responses are averaged over all participants in Experiment 1. Examples are shown for three viewing conditions: 100% contrast at 45 cm (A), 7.5% contrast at 45 cm (B), and 100% contrast at 90 cm (C).

the prior for motion is always centered on 0 in both  $x'$  and  $z'$ . That is, although the sampling distribution of the MAP extends into reversed directions (Figure 7A), the average of this distribution is always in the same direction as the stimulus. Prior studies directly comparing a Bayesian ideal observer to 3-D motion perception either did not present both approaching and receding motion or disregarded direction confusions (Lages, 2006; Welchman et al., 2008), and thus this additional bias was not observed. However, there are several ways in which existing Bayesian models, including the one presented here, may be elaborated to account for these effects. For example, extensions to our model might incorporate a prior that is not centered on zero motion for some stimuli, a cost function that reflects the different behavioral consequences of misperceiving approaching and receding motion, or the impact of attentional effects. Of particular interest would be the exploration of a statistical relationship between stimulus contrast and motion direction in natural scenes.

## Discussion

We have presented a Bayesian model of 3-D motion perception that predicts systematic errors in perceived motion direction, including a lateral bias, a tendency to report approaching motion as receding and vice versa, and a dependency of these errors on viewing distance, contrast, and eccentricity. We tested these predictions in a VR environment where monocular and binocular cues to 3-D motion were available and established that the errors persist under such conditions. Thus, our results demonstrate that uncertainty in retinal-velocity signals, coupled with a prior for slow motion and simple geometric considerations, accounts for a number of motion-perception phenomena in the three-dimensional world. Finally, we identify a limitation of our model: It does not explain why observers report the majority of stimuli as either receding (Figure 9B) or approaching (Figure 9C). Our model provides a framework with which to understand errors in 3-D motion perception at arbitrary locations, and further supports the idea that visual perception can be accurately and parsimoniously modeled as a process of probabilistic inference.

## Previous Bayesian models of motion perception

This work extends a line of Bayesian models that account for errors in motion perception for stimuli presented in the fronto-parallel plane (Hürlimann et al., 2002; Stocker & Simoncelli, 2006; Weiss et al., 2002; Yuille & Grzywacz, 1988). Critically, these models make the assumption that motion percepts reflect noisy measurements of visual motion combined with a prior for slow speeds.

Why would observers employ a prior for slow speeds in the world? A slow-motion prior presumably reflects the fact that objects in the world are most likely stationary, and if moving, are more likely to move slowly rather than quickly. This prior would thus have to disregard the contributions of eye, head, and body motion to the visual input. Nonetheless, even during head-free fixation, it has been shown that retinal-velocity signals are biased towards slower speeds (Aytekin, Victor, & Rucci, 2014). Thus, there is both strong theoretical and experimental evidence that a slow-motion prior is consistent with the statistical regularities of visual experience.

Two groups have previously extended Bayesian motion-perception models to account for errors in the perception of 3-D motion based on binocular cues (Lages, 2006; Lages et al., 2013; Welchman et al., 2008). The model proposed by Welchman and colleagues provides an account for the lateral bias and predicts an effect of viewing distance. However, since the derivation relies on the small-angle approximation, this model does

not account for motion occurring off of the midsagittal plane.

While the Welchman model relied on retinal-velocity cues, the model proposed by Lages and colleagues considered the separate contributions of two binocular cues to 3-D motion—interocular velocity differences and changing binocular disparity signals. The Lages (2006) study concluded that disparity rather than velocity processing introduced the lateral bias. However, the Lages model assumed that prior assumptions operate in retinal coordinates (Lages, 2006; Lages & Heron, 2008). Here we have assumed that the combination of the prior and the likelihood takes place in world coordinates. While these assumptions are essentially equivalent for predicting percepts in the fronto-parallel plane, they can produce different predictions for motion in depth, depending on whether binocular disparity or binocular motion cues are assumed to be the key visual cue (Lages, 2006; Lages et al., 2013). In particular, a Bayesian model based on motion cues being combined in retinal coordinates does not predict a lateral bias in 3-D motion perception.

What is the natural coordinate system in which to formulate a Bayesian perceptual model? We would argue that performing inference in world coordinates makes the most sense because it is ultimately motion in the world and not motion on the retina that is relevant to behavior. Thus, the inference problem should consider a posterior distribution that reflects uncertainty about motion in the world, not motion on the retina. However, the extension of a prior for slow speeds into a probability distribution over 3-D velocities does not have a single solution. For the current model, we assumed (as has been done previously) that the prior distribution (as well as the likelihood and posterior) is represented in a Cartesian world space over  $x'$  and  $z'$ , where motions toward/away and left/right are continuous with each other (i.e., positive and negative arms of the same axis; see heat maps in Figure 7A and 7B). This type of coordinate system is necessary in order for the model to predict the prevalence of direction confusions in depth, because the resulting posterior distribution often straddles  $z' = 0$  but not  $x' = 0$ . From a purely computational perspective, it would be reasonable to consider that the probabilities of motion trajectories might be represented in terms of polar direction and speed. But in such a coordinate system, it is unclear if the same pattern of direction confusions would result. The clear match between the direction-confusion predictions of our model and the experimental data provide strong support that the current model captures essential features that describe the inferences that underlie motion perception.

An additional contribution of the current work is to derive the full sampling distribution of the MAP, and therefore account for trial-to-trial variability in judgments of motion direction. Examination of this

sampling distribution reveals that the Bayesian model accurately predicts that observers will systematically misreport direction of motion in depth such that they judge approaching motion as receding, and vice versa (Fulvio et al., 2015).

### Other 3-D motion cues

While our model is based solely on binocular retinal velocities, the stimuli used in our perceptual experiments included an additional cue to 3-D motion: looming. Consistent visual looming was incorporated in order to mitigate the effect of cue conflicts. Specifically, conflicting cues that indicate zero motion in depth (i.e., no change in retinal size) might lead to an overrepresentation of perceptual errors that are simply due to the cue conflict and are not a feature of 3-D motion perception more generally.

Given the short stimulus duration in our experiments and the relatively large viewing distances, the actual change in stimulus retinal size was minimal on most trials. However, it is important to note that the general presence of looming (and other depth cues such as motion parallax, familiar size, occlusion, etc.) during natural vision may greatly reduce the overall 3-D motion uncertainty. That is, the noise properties plotted in Figure 3 refer to uncertainty from binocular motion alone and do not take any other cues into consideration. At the same time, analyses of motor responses to looming stimuli suggest that other 3-D motion cues can be subject to their own systematic biases based on prior assumptions about the world (López-Moliner, Field, & Wann, 2007; López-Moliner & Keil, 2012). Clearly, future research should work toward unifying the existing Bayesian frameworks for considering different 3-D motion cues.

The perceptual errors we predicted are entirely based on sensory signals produced by object motion, not self-motion. That is, we assumed that observers were stationary and maintained fixation during stimulus presentation. The consequences of self-motion on perceptual accuracy are not directly obvious. While self-motion might increase perceptual errors (Freeman et al., 2010), it also provides additional parallax cues that may help reduce these errors (Aytekin et al., 2014). Ultimately, 3-D motion-perception models will need to incorporate both stimulus and self-motion to help us understand the perceptual accuracy of active observers moving through the real world.

### 3-D motion response paradigms

It is worth highlighting that the average lateral bias reported for the current experiments is considerably larger than the levels of lateral bias reported in previous

work (Duke & Rushton, 2012; Gray, Regan, Castaneda, & Sieffert, 2006; Harris & Dean, 2003; Harris & Drga, 2005; Poljac et al., 2006; Rushton & Duke, 2007; Welchman et al., 2004; Welchman et al., 2008). As mentioned previously, an important distinction between the current experiments and previous studies is that the full 360° space was utilized for both the stimuli and responses. In these previous studies, the stimuli and responses were restricted to approaching motion. An exception is the study conducted by Lages (2006); however, in the data analysis for that experiment, misreports of depth direction were treated as indications that participants were unable to do the task (if occurring on a large proportion of trials) or as “bad” trials in which participants did not see the stimulus. Therefore, such misreports were not treated as a meaningful feature of visual motion perception. Given the motion-in-depth direction confusions observed in the current experiments, many of the measured angular errors between stimulus and response were actually very large, which in turn increases the average lateral bias. To examine this further, in our recent work (Fulvio et al., 2015) we conducted an analysis in which the data were restricted to the range of stimulus motions used in previous studies. Under such conditions, we found a lateral bias that is comparable to the reports of previous studies.

In the current experiments, we focused on response paradigms in which participants indicate the perceived motion direction using a “3-D Pong” setup. Equally interesting would be to examine perceived speed under these viewing conditions (Welchman et al., 2008). Future work could perhaps extend the 3-D Pong set up into a real-time interception task, and examine response times as well as accuracy.

## Errors in the real world

The errors predicted by the current model will no doubt be most apparent in the real world under demanding conditions, such as when there is limited time or poor visibility (Pretto et al., 2012; Shrivastava et al., 2010; Snowden et al., 1998). In situations where sensory uncertainty is very low, the model predicts that these perceptual errors will be negligible. It is difficult to quantify what level of sensory uncertainty a person will be subject to at any particular time during day-to-day life under natural viewing conditions. However, we do know that when stimulus contrast is very high (greater than 100% Weber contrast), the lateral bias can effectively disappear for practiced observers in our experimental setups (Fulvio et al., 2015). While the motion-in-depth confusions persist longer in the laboratory, we expect that these may be similarly reduced by the presence of additional and more reliable visual cues. In fact, the presence of these systematic

errors may provide a way to compare and quantify the performance of different VR display systems, especially those that incorporate less well-understood cues such as predictive head motion or defocus blur.

## Implications for neural processing of motion

While the current model is perceptual and not mechanistic, our predictions and results are relevant to investigating the neural mechanisms that underlie motion perception. The central role of area MT in the processing of binocular 3-D motion signals is now well established, based on both neuroimaging (Rokers, Cormack, & Huk, 2009) and electrophysiology studies (Baker & Bair, 2016; Czuba, Huk, Cormack, & Kohn, 2014; Sanada & DeAngelis, 2014). Our model highlights the fact that both position and binocular speed tuning are essential for inferring the trajectory of a stimulus moving in three dimensions. Consider the case of an object moving directly toward the midpoint between the two eyes. If this object is located in the midsagittal plane, it will cast equal and opposite horizontal velocities in the two eyes. However, if this object has an eccentric location to the left of the midsagittal plane, the velocities cast on the two eyes will not be equal and opposite—they will have opposite signs, but the velocity in the left eye will be greater. Thus, the interpretation of an MT neuron’s tuning profile and preference for 3-D motion must somehow take into account the location of the stimulus relative to the observer, independent of retinotopic location.

When it comes to the slow-motion prior, there remains significant debate on how prior assumptions for visual motion factor into the neural computations, and where perceptual biases arise along the visual motion-processing pathway. Results from neuroimaging show that responses to 2-D motion stimuli can depend on perceived rather than presented speed as early as V1, suggesting that motion priors interact with sensory evidence at the earliest stage of cortical processing (Kok, Brouwer, van Gerven, & de Lange, 2013; Vintch & Gardner, 2014). However, evidence from electrophysiology has been decidedly more mixed (Krekelberg, van Wezel, & Albright, 2006; Livingstone & Conway, 2007; Pack, Hunter, & Born, 2005). Since the biases for the lateral motion and motion-in-depth components for 3-D stimuli have different magnitudes, these differences provide an additional signature for determining whether the responses of particular neuronal populations are driven by the stimulus or the percept.

## Conclusion

Understanding how Bayesian inference plays out during natural vision and natural behavior requires

not only characterizing the prior assumptions of an observer but also having a deep understanding of the sensory signals available at a given point in time. The current model predicts perceived 3-D motion under a wide range of scenarios and viewing conditions, and in doing so provides a parsimonious account of multiple, seemingly disparate perceptual errors.

*Keywords:* motion-3D, motion perception, computational modeling, virtual reality

## Acknowledgments

The authors would like to thank Padmadevan Chettiar for technical support, Michelle Wang and Darwin Romulus for assistance with data collection, and Joe Austerweil for comments on a previous version of this manuscript. EAC was supported by Oculus, Microsoft, and Samsung. BR and JMF were supported by Google. JWP was supported by grants from the Sloan Foundation and the McKnight Foundation, and the NSF CAREER Award (IIS-1150186).

Commercial relationships: none.

Corresponding author: Bas Rokers.

Email: rokers@wisc.edu.

Address: Department of Psychology, University of Wisconsin, Madison, WI, USA.

## References

- Aytekin, M., Victor, J. D., & Rucci, M. (2014). The visual input to the retina during natural head-free fixation. *The Journal of Neuroscience*, *34*(38), 12701–12715.
- Baker, P. M., & Bair, W. (2016). A model of binocular motion integration in MT neurons. *The Journal of Neuroscience*, *36*(24), 6563–6582.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.
- Cooper, E. A., van Ginkel, M., & Rokers, B. (2016). Sensitivity and bias in the discrimination of two-dimensional and three-dimensional motion direction. *Journal of Vision*, *16*(10):5, 1–11, <https://doi.org/10.1167/16.10.5>. [PubMed] [Article]
- Czuba, T. B., Huk, A. C., Cormack, L. K., & Kohn, A. (2014). Area MT encodes three-dimensional motion. *The Journal of Neuroscience*, *34*(47), 15522–15533.
- Dosher, B. A., Sperling, G., & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research*, *26*, 973–990.
- Duke, P. A., & Rushton, S. K. (2012). How we perceive the trajectory of an approaching object. *Journal of Vision*, *12*(3):9, 1–13, <https://doi.org/10.1167/12.3.9>. [PubMed] [Article]
- Farnè, M. (1977). Brightness as an indicator to distance: Relative brightness per se or contrast with the background? *Perception*, *6*, 287–293.
- Freeman, T. C. A., Champion, R. A., & Warren, P. A. (2010). A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Current Biology*, *20*, 757–762.
- Fulvio, J. M., Rosen, M. L., & Rokers, B. (2015). Sensory uncertainty leads to systematic misperception of the direction of motion in depth. *Attention, Perception & Psychophysics*, *77*(5), 1685–1696.
- Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932.
- Gray, R., Regan, D., Castaneda, B., & Sieffert, R. (2006). Role of feedback in the accuracy of perceived direction of motion-in-depth and control of interceptive action. *Vision Research*, *46*(10), 1676–1694, <https://doi.org/10.1016/j.visres.2005.07.036>.
- Harris, J. M., & Dean, P. J. A. (2003). Accuracy and precision of binocular 3-D motion perception. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(5), 869–881.
- Harris, J. M., & Drga, V. F. (2005). Using visual direction in three-dimensional motion perception. *Nature Neuroscience*, *8*(2), 229–233.
- Hürlimann, F., Kiper, D. C., & Carandini, M. (2002). Testing the Bayesian model of perceived speed. *Vision Research*, *42*(19), 2253–2257.
- Jammalamadaka, S. R., & Sengupta, A. (2001). *Topics in circular statistics*. Singapore: World Scientific.
- Johnston, A., & Wright, M. J. (1986). Matching velocity in central and peripheral vision. *Vision Research*, *26*(7), 1099–1109.
- Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology*, *13*(2), 150–158.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1–16.
- Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision*, *7*(7):5, 1–24, <https://doi.org/10.1167/7.7.5>. [PubMed] [Article]

- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719.
- Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge, UK: Cambridge University Press.
- Kok, P., Brouwer, G. J., van Gerven, M. A., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *The Journal of Neuroscience*, *33*(41), 16275–16284.
- Krekelberg, B., van Wezel, R. J., & Albright, T. D. (2006). Adaptation in macaque MT reduces perceived speed and improves speed discrimination. *Journal of Neurophysiology*, *95*(1), 255–270.
- Lages, M. (2006). Bayesian models of binocular 3-D motion perception. *Journal of Vision*, *6*(4):14, 508–522, <https://doi.org/10.1167/6.4.14>. [PubMed] [Article]
- Lages, M. (2013). Straight or curved? From deterministic to probabilistic models of 3D motion perception. *Frontiers in Behavioral Neuroscience*, *7*(79), 1–3.
- Lages, M., & Heron, S. (2008). Motion and disparity processing informs Bayesian 3D motion estimation. *Proceedings of the National Academy of Sciences, USA*, *105*(51), E117.
- Lages, M., Heron, S., & Wang, H. (2013). Local constraints for the perception of binocular 3D motion. In M. Pomplun, & J. Suzuki (Eds.), *Developing and applying biologically-inspired vision systems: Interdisciplinary concepts* (pp. 90–120). New York: IGI Global.
- Levi, D. M., Klein, S. A., & Aitsebaomo, P. (1984). Detection and discrimination of the direction of motion in central and peripheral vision of normal and amblyopic observers. *Vision Research*, *24*(8), 789–800.
- Lin, J. Y., Franconeri, S., & Enns, J. T. (2008). Objects on a collision path with the observer demand attention. *Psychological Science*, *19*(7), 686–692.
- Lin, J. Y., Murray, S. O., & Boynton, G. M. (2009). Capture of attention to threatening stimuli without perceptual awareness. *Current Biology*, *19*(13), 1118–1122.
- Livingstone, M. S., & Conway, B. R. (2007). Contrast affects speed tuning, space-time slant, and receptive-field organization of simple cells in macaque V1. *Journal of Neurophysiology*, *97*(1), 849–857.
- López-Moliner, J., Field, D. T., & Wann, J. P. (2007). Interceptive timing: Prior knowledge matters. *Journal of Vision*, *7*(13):11, 1–8, <https://doi.org/10.1167/7.13.11>. [PubMed] [Article]
- López-Moliner, J., & Keil, M. S. (2012). People favour imperfect catching by assuming a stable world. *PLoS One*, *7*(4), e35705.
- McKee, S. P., Silverman, G. H., & Nakayama, K. (1986). Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Research*, *26*(4), 609–619.
- O’Shea, R. P., Blackburn, S. G., & Ono, H. (1994). Contrast as a depth cue. *Vision Research*, *34*, 1595–1604.
- Pack, C. C., Hunter, J. N., & Born, R. T. (2005). Contrast dependence of suppressive influences in cortical area MT of alert macaque. *Journal of Neurophysiology*, *93*(3), 1809–1815.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.
- Peper, L., Bootsma, R. J., Mestre, D. R., & Bakker, F. C. (1994). Catching balls: How to get the hand to the right place at the right time. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(3), 591–612.
- Poljac, E., Neggers, B., & van den Berg, A. V. (2006). Collision judgment of objects approaching the head. *Experimental Brain Research*, *171*(1), 35–46.
- Pretto, P., Bresciani, J.-P., Rainer, G., & Bühlhoff, H. H. (2012). Foggy perception slows us down. *eLife*, *1*, e00031.
- Rokers, B., Cormack, L. K., & Huk, A. C. (2009). Disparity- and velocity-based signals for three-dimensional motion perception in human MT+. *Nature Neuroscience*, *12*(8), 1050–1055.
- Rushton, S. K., & Duke, P. A. (2007). The use of direction and distance information in the perception of approach trajectory. *Vision Research*, *47*(7), 899–912.
- Sanada, T. M., & DeAngelis, G. C. (2014). Neural representation of motion-in-depth in area MT. *The Journal of Neuroscience*, *34*(47), 15508–15521.
- Schiff, W., Caviness, J. A., & Gibson, J. J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of “looming.” *Science*, *136*(3520), 982–983.
- Schwartz, B. J., & Sperling, G. (1983). Luminance controls the perceived 3-D structure of dynamic 2-D displays. *Bulletin of the Psychonomic Society*, *21*, 456–458.
- Shrivastava, A., Hayhoe, M. M., Pelz, J. B., & Mruczek, R. (2010). Influence of optic flow field restrictions and fog on perception of speed in a virtual driving environment. *Journal of Vision*, *5*(8): 139, <https://doi.org/10.1167/5.8.139>. [Abstract]

- Snowden, R. J., Stimpson, N., & Ruddle, R. A. (1998, April 2). Speed perception fogs up as visibility drops. *Nature*, *392*(6675), 450.
- Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), 578–585.
- Stone, L. S., & Thompson, P. (1992). Human speed perception is contrast dependent. *Vision Research*, *32*(8), 1535–1549.
- Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research*, *22*(3), 377–380.
- Vintch, B., & Gardner, J. L. (2014). Cortical correlates of human motion perception biases. *The Journal of Neuroscience*, *34*(7), 2592–2604.
- Wang, H., Heron, S., Moreland, J., & Lages, M. (2012). A Bayesian approach to the aperture problem of 3D motion perception. *2012 International Conference on 3D Imaging*, 1–8.
- Wei, X. X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain “anti-Bayesian” percepts. *Nature Neuroscience*, *18*(10), 1509–1517.
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604.
- Welchman, A. E., Lam, J. M., & Bühlhoff, 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.
- Welchman, A. E., Tuck, V. L., & Harris, J. M. (2004). Human observers are biased in judging the angular approach of a projectile. *Vision Research*, *44*(17), 2027–2042.
- Yuille, A. L., & Grzywacz, N. M. (1988, May 5). A computational theory for the perception of coherent visual motion. *Nature*, *333*(6168), 71–74.