Annual Review of Vision Science

Seeing a Three-Dimensional World in Motion: How the Brain Computes Object Motion and Depth During Self-Motion

Zhe-Xin Xu[1,2] and Gregory C. DeAngelis[1]

[1]Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York, USA; email: gdeangelis@ur.rochester.edu

[2]Department of Neurobiology, Harvard Medical School, Boston, Massachusetts, USA

Annu. Rev. Vis. Sci. 2025. 11:423–46

First published as a Review in Advance on June 18, 2025

The *Annual Review of Vision Science* is online at vision.annualreviews.org

https://doi.org/10.1146/annurev-vision-110323-112124

Keywords

motion perception, optic flow, depth perception, motion parallax, visual cortex, neural computation

Abstract

Humans and other animals move their eyes, heads, and bodies to interact with their surroundings. While essential for survival, these movements produce additional sensory signals that complicate visual scene analysis. However, these self-generated visual signals offer valuable information about self-motion and the three-dimensional structure of the environment. In this review, we examine recent advances in understanding depth and motion perception during self-motion, along with the underlying neural mechanisms. We also propose a comprehensive framework that integrates various visual phenomena, including optic flow parsing, depth from motion parallax, and coordinate transformation. The studies reviewed here begin to provide a more complete picture of how the visual system carries out a set of complex computations to jointly infer object motion, self-motion, and depth.

1. INTRODUCTION

Visual image motion provides a powerful source of information for many important visual computations (Albright & Stoner 1995, Gibson 1950, Johansson 1975, Lu & Sperling 1995, Nishida et al. 2018, Park & Tadin 2018, Shadlen & Newsome 1996). For example, the image motion of objects enables us to discern how those objects move in the world (Park & Tadin 2018, Stone et al. 2000), global patterns of visual motion (i.e., optic flow) allow us to infer our self-motion (i.e., translation and rotation) through the world (Britten 2008, Gibson 1950, Koenderink 1986, Lappe et al. 1999), and relative visual motion (i.e., motion parallax) provides a powerful cue to 3D scene structure (Andersen & Bradley 1998; Bradshaw & Rogers 1996; Gibson et al. 1959; Gogel & Tietz 1979; Koenderink & van Doorn 1991; Nawrot 2003; Nawrot & Blake 1989; Nawrot & Joyce 2006; Nawrot & Stroyan 2009; Ono et al. 1986, 1988; Rogers 1993, 2009; Rogers & Graham 1979, 1982; Smith & Smith 1963; Stroyan & Nawrot 2012; Ullman 1979). Many of our most fundamental visually guided behaviors, such as reaching toward and grasping an object, require the visual system to estimate the motion of objects in the world (i.e., scene-relative motion) as well as their 3D location in space. The ease with which we perform these tasks belies the difficulty of the underlying neural computations. Understanding the visual processing of object motion and depth during self-motion also has important applications to emerging technologies, such as self-driving vehicles, robotic systems, and virtual or augmented reality.

When the eye is stationary relative to the scene, there is a direct mapping from scene-relative object motion to retinal image motion. Indeed, much classic vision science at both the behavioral and neurophysiological levels has been performed under experimental conditions in which the eyes and head remain fixated during stimulus presentation. While a great deal has been learned under these restricted conditions, the problem is substantially more complex when the eye translates and rotates relative to the scene as a consequence of eye, head, and body movements. In general, the retinal velocity of an object is determined by three variables: how the object moves relative to the scene, how the eye moves relative to the scene (i.e., self-motion), and where the object is located in 3D space (i.e., depth), as illustrated in **Figure 1*a***. If these three variables are specified, then retinal velocity can be calculated via projective geometry (**Figure 1*b***). However, the inverse (i.e., inference) process is much more difficult, as there can be many possible combinations of object motion, self-motion, and depth that produce the same retinal velocity. Thus, crucially, to estimate any of these three variables accurately, the brain needs to infer the respective contributions of each variable to retinal velocity. For example, to compute an object's scene-relative motion, the brain needs to discount image motion resulting from self-motion (Dokka et al. 2015b; Foulkes et al. 2013; MacNeilage et al. 2012; Niehorster & Li 2017; Peltier et al. 2020; Royden & Connors 2010; Rushton et al. 2018; Warren & Rushton 2007, 2008, 2009a,b). To compute an object's depth from motion parallax (Bradshaw & Rogers 1996; Nawrot 2003; Nawrot & Joyce 2006; Nawrot & Stroyan 2009; Ono et al. 1988; Rogers & Graham 1979, 1982; Rogers & Rogers 1992; Stroyan & Nawrot 2012), the brain needs to discount image motion resulting from scene-relative object motion and isolate the component of image motion associated with self-motion (French & DeAngelis 2020, 2022).

How does the brain parse image motion into contributions from object motion, self-motion, and depth? How does the brain compensate for self-motion to compute scene-relative object motion, or vice versa? How does the brain estimate depth from motion parallax when objects are moving in the world? How does the brain flexibly compensate for the visual consequences of smooth eye movements under different viewing geometries? The goal of this review is to summarize recent progress in addressing these types of questions at both the behavioral and neurophysiological levels. While behavioral studies over the past decade or two have made substantial

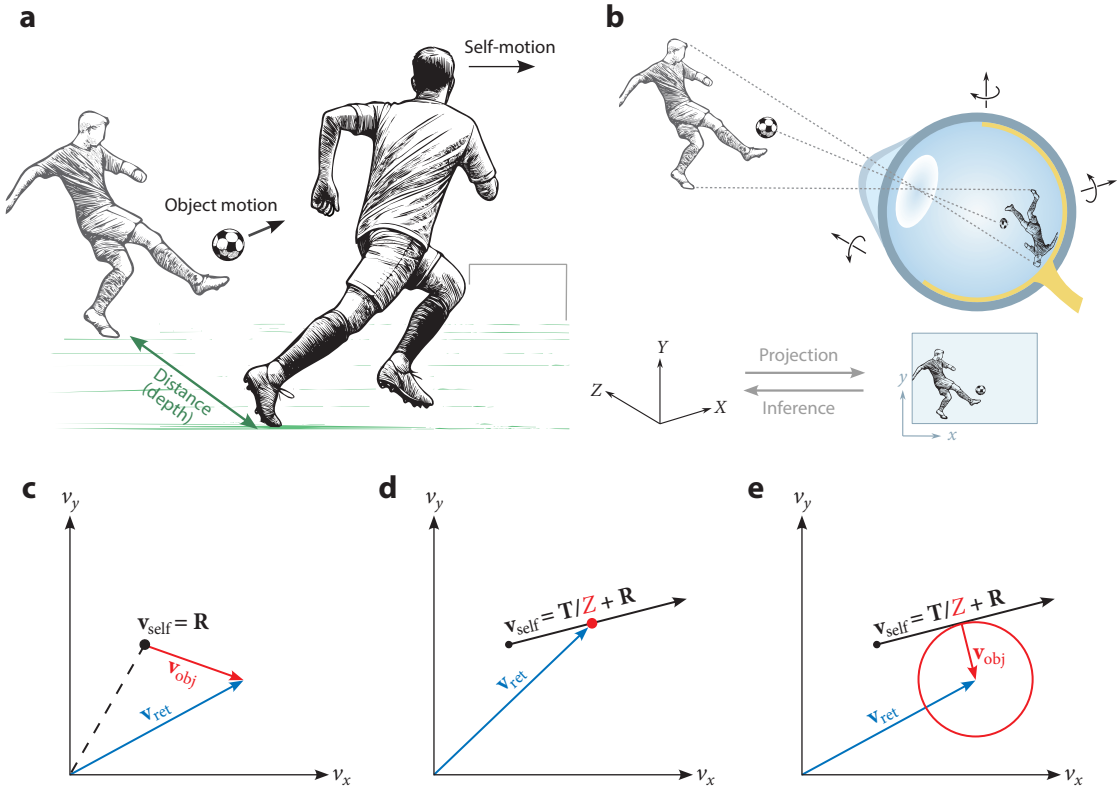

Figure 1

Interactions between self-motion, object motion, and depth, as well as the constraint ray approach to visualizing their relationships. (*a*) The retinal image motion of an object, such as a soccer ball, depends on the three-way interactions between its motion in the world, its distance (or depth) to the observer, and the observer's self-motion. (*b*) An observer's self-motion can be described in a 3D Cartesian coordinate system, *XYZ*, with rotation and translation along each axis (*black arrows*). Objects in the scene, such as the opponent and the soccer ball, project onto a 2D retinal coordinate system, *xy*. The retinal image motion of an object can be computed based on the projective geometry from the 3D world coordinate system, *XYZ*, to the 2D image coordinate system, *xy*. (*c–e*) Illustration of the constraint ray approach to describing self-motion and object motion in a 2D image velocity space. (*c*) When only rotational self-motion is present, the optic flow vector at any given point on the retina is defined (*black dot*). When retinal image motion (*blue vector*) deviates from the optic flow vector, their difference is explained as object motion (*red vector*). (*d*) When both translational and rotational self-motion components are present and the object's depth, *Z*, is not known, a stationary object's velocity is constrained to lie anywhere along a ray (*black vector*). The black dot indicates the flow vector when $Z \to \infty$. A given point on the ray corresponds to a particular depth, *Z*. If the retinal motion of an object (*blue vector*) falls on the ray, it can be fully explained by self-motion. Therefore, the object should be perceived as stationary and located at the depth, *Z*, given by the self-motion constraint (*red dot*). (*e*) If the retinal motion of an object deviates from the constraint ray, the difference might be perceived as independent object motion relative to the scene (*red vector*). The soccer player images in panels *a* and *b* were created with ChatGPT-4o/DALL-E.

progress in addressing some of these questions, there is still much to learn about the neural mechanisms underlying the visual system's remarkable capacity to make joint inferences over object motion, self-motion, and depth. In addition to reviewing recent progress, we also describe how visual computations related to object motion and depth can be unified under a computational framework that considers viewing geometry and the constraints imposed by self-motion.

Few studies to date have examined the full three-way interactions between object motion, self-motion, and depth. Thus, we initially review studies that examine interactions between object motion and self-motion and subsequently review studies that examine how scene-relative object

motion interacts with computations of depth. Finally, we consider a framework that bridges some previously separate lines of work.

2. OBJECT MOTION PROCESSING DURING SELF-MOTION

In this section, we explore mechanisms through which the brain estimates object motion while the observer is moving. During self-motion, retinal image motion, $\mathbf{v}_{\text{ret}}$, has two main components: one that reflects scene-relative object motion, $\mathbf{v}_{\text{obj}}$, and one that arises from translation and rotation of the eye relative to the scene (i.e., self-motion), $\mathbf{v}_{\text{self}}$:

$$\mathbf{v}_{\text{ret}} = \mathbf{v}_{\text{obj}} + \mathbf{v}_{\text{self}}. \quad 1.$$

Both of these components of retinal image motion also generally depend on the distance of a point in the scene from the observer (i.e., depth). To estimate scene-relative object motion, the brain needs to estimate and discount $\mathbf{v}_{\text{self}}$. In general, three types of information contribute to the solution of this problem: optic flow, depth cues, and extraretinal signals. Optic flow provides information about how the eye translates and rotates relative to the scene. Depth cues, such as binocular disparity and relative size, can help disambiguate sources of retinal image motion by providing information about 3D scene structure. Extraretinal signals, including vestibular signals related to head movements and efference copy of motor commands that generate eye movements, provide independent sources of information about self-motion that can help tease apart contributions from object motion and self-motion. This section reviews literature regarding how each of these sources of information contributes to estimating object motion during self-motion (and vice versa).

2.1. Optic Flow and the Constraint Ray of Self-Motion

Before reviewing empirical studies, we summarize a theoretical framework that helps tie together various empirical observations: the constraint ray of self-motion (Nelson 1991, Thompson & Pong 1990). Optic flow refers to large-field background image motion that is produced by self-motion, specifically the translation and rotation of the eye relative to a scene (Gibson 1950). At a given point (x, y) on the retina, the optic flow of an object located at depth Z is related to self-motion with six degrees of freedom, including translation of the eye along the X-, Y-, and Z-axes (T_X, T_Y, and T_Z) and rotation of the eye about these axes (R_X, R_Y, and R_Z; see **Figure 1*b***) (Longuet-Higgins & Prazdny 1980):

$$\mathbf{v}_{\text{self}} = \mathbf{T}/Z + \mathbf{R}, \quad 2.$$

where $\mathbf{v}_{\text{self}}$ represents the 2D optic flow vector at a point in the visual field,

$$\mathbf{v}_{\text{self}} = \begin{pmatrix} v^x_{\text{self}} \\ v^y_{\text{self}} \end{pmatrix},$$

$\mathbf{T}$ denotes the image motion generated by translational self-motion,

$$\mathbf{T} = \begin{pmatrix} -T_X + xT_Z \\ -T_Y + yT_Z \end{pmatrix},$$

and $\mathbf{R}$ represents the image motion produced by rotational self-motion,

$$\mathbf{R} = \begin{pmatrix} xyR_X - (x^2 + 1) R_Y + yR_Z \\ (y^2 + 1) R_X - xyR_Y - xR_Z \end{pmatrix}.$$

Many algorithms have been developed to compute optic flow from image sequences (Barron et al. 1994, Horn & Schunck 1981, Lucas & Kanade 1981), and it is well established that humans and monkeys can infer their direction of heading from optic flow (Britten 2008, Fetsch et al. 2009, Gibson 1950, Gu et al. 2008, Koenderink 1986, Lappe et al. 1999, Warren & Hannon 1988).

Equation 2 imposes an important constraint on the interpretation of retinal motion—any image motion that deviates from this relationship can be attributed to scene-relative object motion (Jain 1984, Nelson 1991, Thompson & Pong 1990). This can be visualized as a constraint ray in a 2D image velocity space (**Figure 1*c*–*e***). Here, the *x*- and *y*-axes represent the horizontal and vertical components of retinal image velocity at a particular location on the retina, respectively, such that a point on the 2D velocity plane represents a particular image velocity at a location, (x, y). When only rotational self-motion is present (i.e., pure eye rotation relative to the scene),[1] the optic flow vector at any given point on the retina is depth invariant and well defined (Equation 2, $\mathbf{T} = \mathbf{0}$; **Figure 1*c***, black dot). When retinal image motion deviates from the optic flow vector, their difference can be explained as scene-relative object motion (**Figure 1*c***, red vector). When both translational and rotational components of self-motion are present, all possible flow vectors can be visualized as a ray in this 2D plane, with location along the ray being scaled by distance, Z (**Figure 1*d***). When the object is sufficiently far away from the observer (i.e., $Z \to \infty$), the translation component in Equation 2, $\mathbf{T}/Z$, approaches zero, and the optic flow vector is solely determined by the rotation component, $\mathbf{R}$ (**Figure 1*d***, black dot). As the object gets closer to the observer, Z decreases, and the translational component of the flow vector scales up in magnitude according to Equation 2. Thus, this constraint forms a ray in the 2D image velocity plane, and a given point on the ray corresponds to a particular depth of a stationary object in the scene.

When the retinal image motion of an object deviates from the constraint ray, object motion can be detected (**Figure 1*c*,*e***), and the direction of object motion can be determined by subtracting $\mathbf{v}_{self}$ from $\mathbf{v}_{ret}$. When the object's depth is not well specified (**Figure 1*e***), the exact magnitude and direction of object motion may be ambiguous, and the shortest vector might be taken as $\mathbf{v}_{obj}$ due to a slow speed prior (Stocker & Simoncelli 2006). Royden & Connors (2010) found that humans are better at detecting object motion when its direction deviates more from the constraint ray, implying the use of self-motion constraints in radial optic flow. This has been tested more directly and confirmed in a recent study by Lutwak et al. (2025).

2.2. Optic Flow Parsing

When an apple falls from a tree, we expect it to fall vertically due to the force of gravity (**Figure 2*a***). However, as an observer walks by the tree, the optic flow produced by the observer's self-motion adds to the apple's motion and changes its image velocity on the retina (**Figure 2*b*,*c***). To estimate the apple's motion in the world, one must subtract the optic flow vector associated with self-motion from the retinal image motion of the apple (**Figure 2*b*,*c***). Optic flow parsing has been proposed as a visual mechanism by which the brain estimates scene-relative object motion in the presence of self-motion (Foulkes et al. 2013; Niehorster & Li 2017; Royden & Connors 2010; Rushton et al. 2018; Warren & Rushton 2007, 2008, 2009a). Formally, the flow parsing hypothesis states that the brain computes object motion in the world, $\mathbf{v}_{obj}$, by subtracting the optic flow vector at the location of the object, $\mathbf{v}_{self}$, from the object's retinal velocity, $\mathbf{v}_{ret}$ (**Figures 1*c*,*e*** and **2*b*,*c***):

$$\mathbf{v}_{obj} = \mathbf{v}_{ret} - \mathbf{v}_{self}. \qquad 3.$$

[1]Pure rotations of the eye relative to the scene are probably less common than most people think. Anytime the head rotates, this introduces some eye translation, since the axis of rotation of the neck is well offset from the axes of rotation of the eyes.

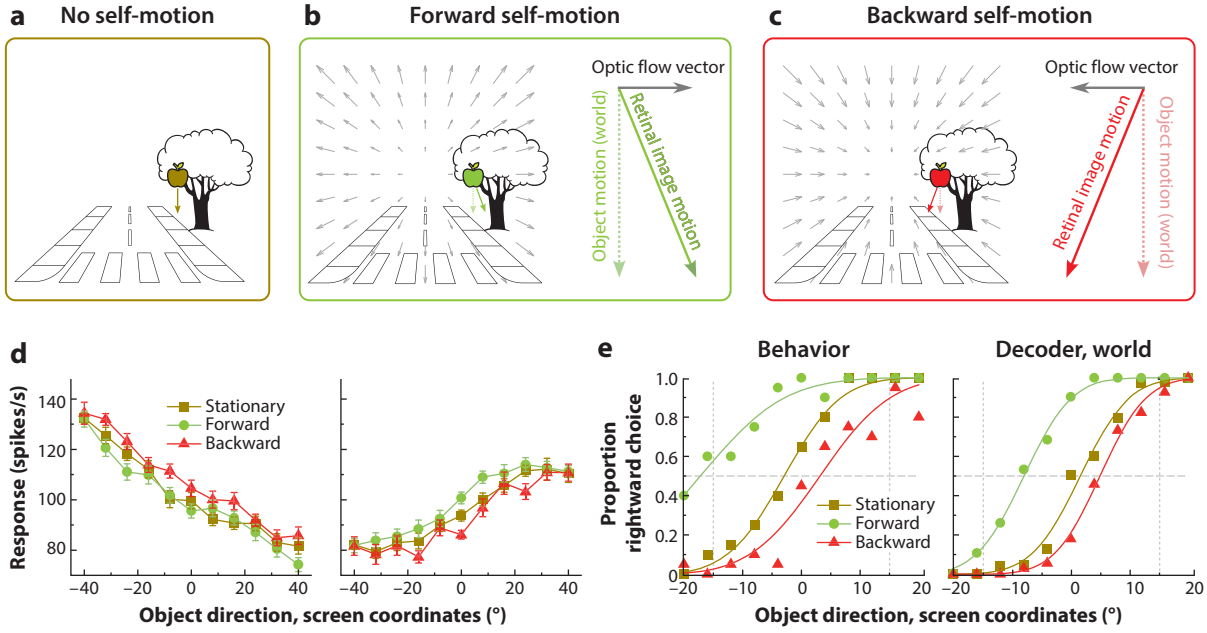

Figure 2

Optic flow parsing and its neural correlates. (*a*) When the observer is stationary, the retinal image motion of an object such as an apple (*gold vector*) reflects its motion in the world. (*b*, *left*) When the observer moves forward, optic flow vectors (*gray*) are added to the object's motion (*dashed green vector*), resulting in an oblique retinal image motion (*solid green vector*). (*Right*) Object motion in the world (*dashed green vector*) can be computed by subtracting the optic flow vector (*gray vector*) from the retinal motion vector (*solid green vector*). (*c*) When the observer moves backward, the optic flow vector is in the opposite direction of that during forward self-motion (*gray vectors*), and the resulting retinal motion is tilted to the left (*solid red vector*). The same subtraction can be performed to infer object motion in world coordinates (*dashed red vector*). (*d*) Example tuning curves of single neurons in the middle temporal (MT) area. Compared to the stationary condition (*gold curve*), tuning curves shifted in opposite directions when forward (*green curve*) or backward (*red curve*) optic flow was present. (*e*, *left*) Psychometric function describing one animal's behavior in a direction discrimination task. (*Right*) Psychometric function from a decoder trained to predict world-centered object motion direction from a small population of MT neurons. Both behavior and decoder performance demonstrate shifts consistent with a partial flow parsing effect in forward (*green*) and backward (*red*) self-motion conditions. Panels *d* and *e* adapted from Peltier et al. (2024) with permission from Elsevier.

Flow parsing predicts that perception of object motion on a visual display will be biased toward directions opposite to the optic flow vector at the location of the object. Indeed, numerous studies have demonstrated that humans and macaques exhibit systematic biases in perceived motion that depend on visual field location and eccentricity in the manner predicted by the flow parsing hypothesis (Foulkes et al. 2013; MacNeilage et al. 2012; Mayer et al. 2021; Niehorster & Li 2017; Peltier et al. 2020; Royden & Connors 2010; Rushton et al. 2018; Warren & Rushton 2007, 2008, 2009a,b). Notably, these effects are still seen (at reduced magnitude) when optic flow is present only in the visual hemifield opposite to the location of a moving object (Peltier et al. 2020, Warren & Rushton 2009a), suggesting a global component to the computation. There is strong evidence that humans and macaques may underestimate $\mathbf{v}_{self}$, leading to imperfect flow parsing (Foulkes et al. 2013; Niehorster & Li 2017; Peltier et al. 2020; Warren & Rushton 2007, 2009a). Thus, a revised model incorporates a flow parsing gain, g, on $\mathbf{v}_{self}$ to capture this underestimation (Niehorster & Li 2017):

$$\mathbf{v}_{obj} = \mathbf{v}_{ret} - g\mathbf{v}_{self}, \quad 4.$$

where $g = 1$ implies perfect flow parsing, $0 \leq g < 1$ indicates an underestimation of $\mathbf{v}_{self}$, and $g > 1$ implies an overestimation. A recent study suggests that humans can estimate self-motion more accurately in an immersive 3D virtual environment, resulting in a flow parsing gain closer to 1 (Jörges & Harris 2022).

A recent electrophysiological study (Peltier et al. 2024) has identified neural correlates of optic flow parsing in the macaque middle temporal (MT) area. Macaques exhibit perceptual biases in line with the flow parsing hypothesis when trained to perform a motion discrimination task in the presence of radial optic flow simulating forward or backward self-motion (Peltier et al. 2020, 2024). The responses of MT neurons are modulated by background optic flow that lies well outside of their receptive fields (**Figure 2*d***), and these modulations depend on the neurons' direction preferences in a manner that is consistent with shifting population activity to account for the behavioral biases (Peltier et al. 2024). Correspondingly, the decoding of small populations of MT neurons predicts perceptual biases similar to those observed behaviorally (**Figure 2*e***), indicating that MT activity can at least partially account for the behavioral effects of flow parsing (Peltier et al. 2024). Modulations of MT responses by optic flow are substantially delayed relative to visual response onset (Peltier et al. 2024), suggesting a potential role for feedback in computing scene-relative object motion during self-motion.

2.3. Contribution of Depth Cues and Vestibular Signals

As described in Section 2.1, humans can detect object motion that deviates from the constraint ray (**Figure 1**). In most cases, this corresponds to situations in which the vector of scene-relative object motion intersects with the constraint ray. However, a particularly challenging scenario arises when objects move in the same direction as the local optic flow vector (**Figure 3*a*,*b***). In such situations, image motion associated with scene-relative object motion may be confounded with motion parallax resulting from self-motion. A solution to this problem is to have access to independent cues regarding the depth of the object (Z in Equation 2), in which case the constraint ray reduces to a specific point along the ray (**Figure 3*c***, blue dot). Consistent with this expectation, Rushton et al. (2007) showed that moving objects automatically pop out in a scene when they are rendered with binocular disparity cues, indicating the contribution of depth cues in disambiguating the source of retinal motion.

Using an experimental design similar to that of Rushton et al. (2007), Kim et al. (2022) trained macaque monkeys to detect scene-relative object motion (**Figure 3*a***). Under conditions in which scene-relative object motion and depth were confounded due to the lack of additional depth cues (as in **Figure 3*a*,*b***), macaques could detect scene-relative object motion only when binocular disparity cues were provided and were in conflict with depth defined by motion parallax (**Figure 3*f***). Previously, MT neurons were shown to encode depth from binocular disparity cues (Born & Bradley 2005, DeAngelis & Newsome 1999, DeAngelis & Uka 2003, Maunsell & Van Essen 1983, Uka & DeAngelis 2003, Zeki 1980) and motion parallax cues (Kim et al. 2015a,b, 2016, 2017; Nadler et al. 2008, 2009, 2013). Moreover, Nadler et al. (2013) identified a subset of MT neurons that show opposite depth tuning preferences for disparity and motion parallax cues, and they speculated that such neurons might play a specific role in detecting scene-relative object motion. Indeed, Kim et al. (2022) found that these MT neurons, referred to as opposite cells, often responded better when a cue conflict between disparity and motion parallax indicated scene-relative object motion (e.g., **Figure 3*e***), and they were selectively correlated with perceptual detection of object motion in the presence of self-motion. In contrast, MT neurons with congruent depth preferences based on disparity and motion parallax typically responded maximally to objects that were stationary in the world (e.g., **Figure 3*d***), and their responses

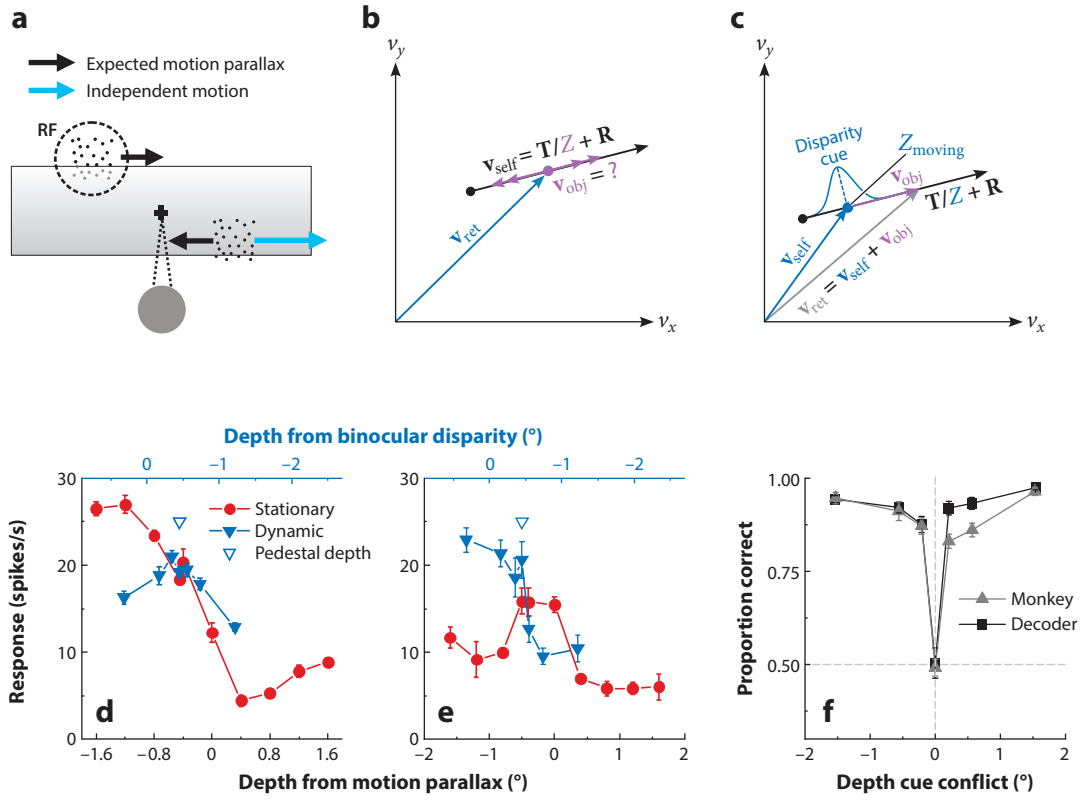

Figure 3

Object motion detection based on cue conflict between motion parallax and binocular disparity cues. (*a*) When an observer translates laterally and fixates on a stationary point in the world (*black plus sign*), the retinal image motion of objects (*black dots*) can come from either motion parallax (*black arrows*) or independent object motion (*cyan arrow*). (*b*) When an object moves in the same direction as the optic flow vector produced by self-motion (*purple vectors*), it becomes challenging to attribute retinal motion to either self-motion or object motion. (*c*) This problem is resolved by a reliable depth cue, such as binocular disparity, which reduces the self-motion constraint to a single point (*blue dot*). (*d*, *e*) Example neurons recorded from area MT in response to moving (i.e., dynamic) and stationary objects. (*d*) Data from a neuron with congruent preferences for near depth based on disparity and motion parallax cues. The red curve shows depth tuning for stationary objects, and the blue curve shows depth tuning for moving (i.e., dynamic) objects. (*e*) Data from a neuron with opposite depth preferences based on disparity and motion parallax, formatted as in panel *d*. Note that responses to some moving objects (*blue*) exceed responses to all stationary objects (*red*). (*f*) Proportion of correct reports as a function of the amount of cue conflict between binocular disparity and motion parallax cues to depth. Gray triangles show behavior from one monkey, and black squares show predicted behavior from a decoder of MT responses. Abbreviations: MT, middle temporal; RF, receptive field. Panels *a* and *d–f* adapted from Kim et al. (2022) (CC BY 4.0).

were not predictive of perceptual detection. Finally, Kim et al. (2022) showed that decoding populations of MT neurons could account for the monkey's performance in the object detection task (**Figure 3*f***). Thus, opposite cells are well suited to identifying situations in which the depth expected from motion parallax is not consistent with the depth expected from binocular disparity, with such deviations being attributable to scene-relative object motion, as illustrated in **Figure 3*c***.

Humans and other animals integrate visual and vestibular signals to achieve a more accurate and precise perception of self-motion (Dokka et al. 2015a, Fetsch et al. 2009, Gu et al. 2008, Morgan et al. 2008). As discussed in Section 2.2, flow parsing gain is often < 1 in studies using purely visual stimuli, indicating that only a portion of self-motion is accounted for (Foulkes et al.

2013; Niehorster & Li 2017; Peltier et al. 2020; Warren & Rushton 2007, 2008, 2009a). Some studies have shown that in the presence of vestibular signals, estimation of self-motion becomes more accurate, leading to a flow parsing gain closer to 1 (Dokka et al. 2015b, MacNeilage et al. 2012, Peltier et al. 2020). These findings indicate the contribution of vestibular signals to object motion processing during self-motion. How multisensory heading perception interacts with optic flow parsing at the cellular level in the visual cortex remains an open question.

2.4. Object Motion Processing During Smooth Pursuit Eye Movements

Eye movements are integral to visual perception. Rapid saccadic eye movements swiftly shift our gaze from one point to another, while smooth pursuit eye movements continuously follow moving objects (Yarbus 2013). In contrast to saccadic movements, which are brief and temporarily suppress visual perception (Matin 1974), smooth pursuit eye movements often persist for extended periods of time, such that it would not be viable to suppress perception during smooth pursuit (Spering & Montagnini 2011). The extended time course of smooth pursuit eye movements introduces a challenge: Smooth pursuit movements add a component of retinal image motion to everything within the visual scene, even stationary objects. How does the visual system form a coherent percept of the world in the face of the sensory effects of self-generated eye movements?

Some of the earliest explorations of motion perception during smooth pursuit include the Aubert–Fleischl phenomenon and the Filehne illusion. The Aubert–Fleischl phenomenon describes an underestimate of a moving object's speed when the eyes track that object by smooth pursuit (Aubert 1887, von Fleischl 1882). Filehne (1922) described an illusion in which stationary objects appear to move during smooth pursuit eye movements. The Filehne illusion and Aubert–Fleischl phenomenon are traditionally thought to reflect a process in which the brain compensates for the visual consequences of smooth pursuit but underestimates pursuit eye velocity (Festinger et al. 1976, Mack & Herman 1973). More generally, others (Wertheim 1981, 1987) have suggested that the brain uses a reference signal, which combines visual and extraretinal signals about eye velocity, to discount image motion associated with smooth pursuit. A Bayesian treatment of this problem has also been proposed (Freeman et al. 2010).

Regardless of the source of signals related to smooth pursuit eye movements, the dominant conceptual model is that the visual system compensates for smooth pursuit through a motion vector computation (Freeman et al. 2010; Souman et al. 2006a,b; Wertheim 1981, 1987). This computation can be understood as a transformation from retinal-centered to world-centered reference frames. This coordinate transformation can be achieved by simply adding the eye velocity, $\mathbf{v}_{eye}$, to the retinal velocity of the object, $\mathbf{v}_{ret}$:

$$\mathbf{v}_{obj} = \mathbf{v}_{ret} + \mathbf{v}_{eye}. \quad 5.$$

While variations of this scheme have been proposed to account for the retinal and extraretinal sources of $\mathbf{v}_{eye}$ (Wertheim 1981) or to explain the differential signal gains (Freeman & Banks 1998), the general form of a vector addition computation between retinal and eye velocities fits well with numerous observations from human psychophysics (Freeman & Banks 1998; Freeman et al. 2010; Souman & Freeman 2008; Souman et al. 2006a,b). However, as we describe in Section 4, this general model for computing object motion during smooth pursuit applies only to the case of pure eye rotations and fails dramatically in the presence of eye translations.

The neural coding of eye position and pursuit velocity has been studied extensively in the cerebral cortex. Early work demonstrated that neurons in parietal areas jointly encode object location on the retina and gaze direction in a multiplicative manner, which is supported by both experimental and theoretical findings (Andersen 1989, Andersen & Mountcastle 1983, Andersen & Zipser 1988, Pouget & Sejnowski 1997, Salinas & Abbott 1995). More recent studies suggest

that this process starts as early as primary visual cortex (V1) (Morris & Krekelberg 2019, Parker et al. 2022b). Morris & Krekelberg (2019) showed that similar gain fields exist in populations of V1 neurons for encoding eye position and retinal images during saccades and smooth pursuit, enabling downstream regions to read out object locations in the world.

In macaques, area MT and the medial superior temporal (MST) area have been shown to make causal contributions to the initiation of smooth pursuit eye movements (Groh et al. 1997; for reviews, see Lisberger 2015, Lisberger et al. 1987). Similarly, coordinate transformations of motion signals have been found in area MST (Brostek et al. 2015; Chukoskie & Movshon 2009; Inaba et al. 2007, 2011) and other parietal areas (Ilg et al. 2004). For example, Chukoskie & Movshon (2009) measured the speed tuning of MT and MST neurons in response to a random-dot patch moving within their receptive fields during either fixation or smooth pursuit behaviors, and they found a fraction of neurons with responses better explained by head-centered speed tuning in both areas (15% of MT cells and 24% of MST cells). Using a large-field random-dot motion stimulus, Inaba et al. (2011) reported that the direction and speed tuning of MST cells are substantially shifted toward head-centered coordinates, while few MT neurons exhibited similar tuning shifts. Brostek et al. (2015) showed that MST neurons encode retinal image motion and eye velocity using a gain field during optokinetic responses. A common theme in these studies is that MT neurons primarily encode retinal motion, whereas MST neurons encode motion in reference frames that shift toward head-centered coordinates. Conventionally, most electrophysiological studies of reference frames have implicitly assumed that neurons encode spatial variables with a fixed reference frame. However, a recent study has demonstrated flexible coding of motion in the parietal cortex, such that populations of neurons can represent object motion in multiple frames of reference depending on task instructions (Sasaki et al. 2020). These findings might explain some variability in outcomes across previous studies of neural reference frames.

Counter to the conventional view that MT neurons encode retinal velocity, other more recent studies have shown that the visual responses of MT neurons can be strongly modulated by smooth pursuit eye movements. For example, MT neurons exhibit depth selectivity from motion parallax cues, and this selectivity was shown to be mediated by smooth eye movement command signals rather than vestibular signals related to head movement (Nadler et al. 2008, 2009). These findings fit well with the motion-pursuit law (Nawrot & Stroyan 2009), which showed that depth can be computed from the combination of motion parallax and scene-relative eye velocity but does not require information about head movement. An important extension of these studies demonstrated that large-field background motion that simulates eye rotation also generates depth-sign selectivity in MT neurons (Kim et al. 2015b). Thus, optic flow induced by smooth eye movements can itself be used as a proxy for extraretinal signals about eye velocity in neural computations, a topic to which we return in Section 4. A recent study showed that neurons in the dorsal part of area MST (MSTd) have congruent selectivity for direction of eye rotation based on both optic flow and extraretinal signals, suggesting MSTd as a possible source of integrated information about eye rotation (DiRisio et al. 2023).

2.5. Causal Inference in the Perception of Object Motion and Self-Motion

So far, we have reviewed how self-motion can affect object motion perception and the roles of various signals in compensating for the visual consequences of self-motion. However, the presence of moving objects in a scene also disrupts the optic flow field produced by self-motion. Because humans automatically infer self-motion from optic flow (Warren & Hannon 1988), object motion can bias self-motion perception (Acerbi et al. 2018; Dokka et al. 2015a, 2019; Fajen & Kim 2002; Layton & Fajen 2016a,b) (**Figure 4**).

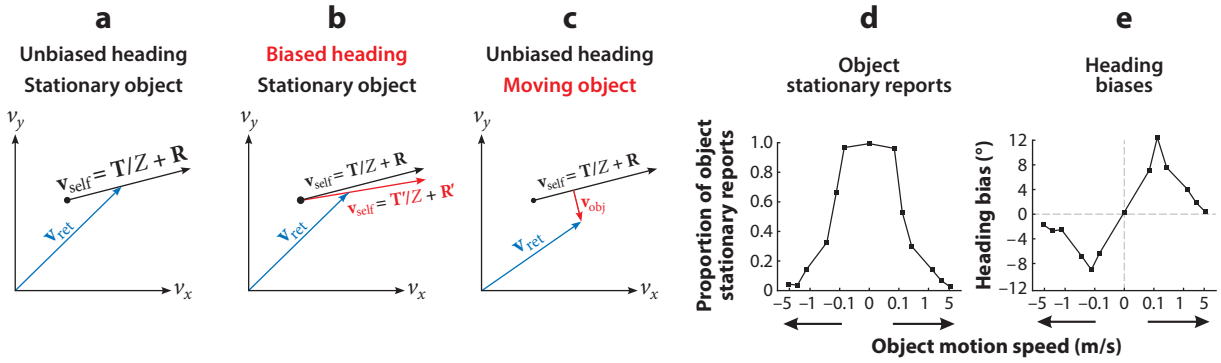

Figure 4

Causal inference in self-motion perception. (*a*) When retinal motion is in line with the constraint ray, no object motion should be perceived. (*b*) When the retinal motion of an object slightly deviates from the constraint ray, the object might still be perceived as stationary in the world, and the slight difference might be used to alter the perception of self-motion, resulting in an updated constraint ray (*red vector*) that is consistent with the retinal motion. (*c*) When the retinal motion of an object sufficiently differs from the optic flow and does not lie along the constraint ray, independent object motion is perceived (*red vector*). (*d*, *e*) Data from an example subject showing stationarity reports (*d*) and heading bias (*e*) as a function of object motion speed. When the object motion speed is zero, the scenario corresponds to panel *a*. When object speed is small, the object is perceived as stationary (*d*), and its motion causes a bias in heading perception (*b* and *e*). When the object speed is sufficiently large, the retinal motion of the object is attributed to object motion (*c*) and heading bias decreases (*e*). Panels *d* and *e* adapted from Dokka et al. (2019).

The interpretation of retinal image motion is fundamentally a causal inference problem (French & DeAngelis 2020, Körding et al. 2007, Shams & Beierholm 2010)—the observer must attribute the source of retinal motion to self-motion, object motion, or both (Acerbi et al. 2018, Dokka et al. 2019, French & DeAngelis 2020). One possible solution to this problem is to examine the difference between retinal motion and the inferred optic flow vector. When retinal motion is consistent with the global optic flow, meaning that $\mathbf{v}_{ret}$ lies along the constraint ray (**Figure 4*a***), its source can be explained by self-motion alone, and one should not infer that there is a moving object. When retinal motion only slightly deviates from the constraint ray, one might still perceive the object as stationary in the world and integrate its retinal motion with the background optic flow to update the perception of self-motion, which results in a new constraint ray that is compatible with the observed retinal image motion (**Figure 4*b***). When retinal motion sufficiently differs from optic flow and does not lie along the constraint ray (**Figure 4*c***), a moving object must be present (Acerbi et al. 2018, Dokka et al. 2019). The scenarios depicted in **Figure 4*a–c*** provide an intuitive understanding of a probabilistic solution formulated under the framework of Bayesian inference, and the theoretical predictions have been shown to match results from human psychophysical experiments reasonably well (**Figure 4*d*,*e***) (Acerbi et al. 2018, Dokka et al. 2019, Körding et al. 2007, Shams & Beierholm 2010).

3. DEPTH FROM MOTION PARALLAX AND INTERACTIONS WITH OBJECT MOTION

In the previous section, we examined how self-motion influences inferences about scene-relative object motion, and vice versa. Another variable that can strongly influence retinal image velocity is depth. In this section, we review how the brain computes depth from motion parallax, and then we examine how scene-relative object motion influences depth perception, and vice versa.

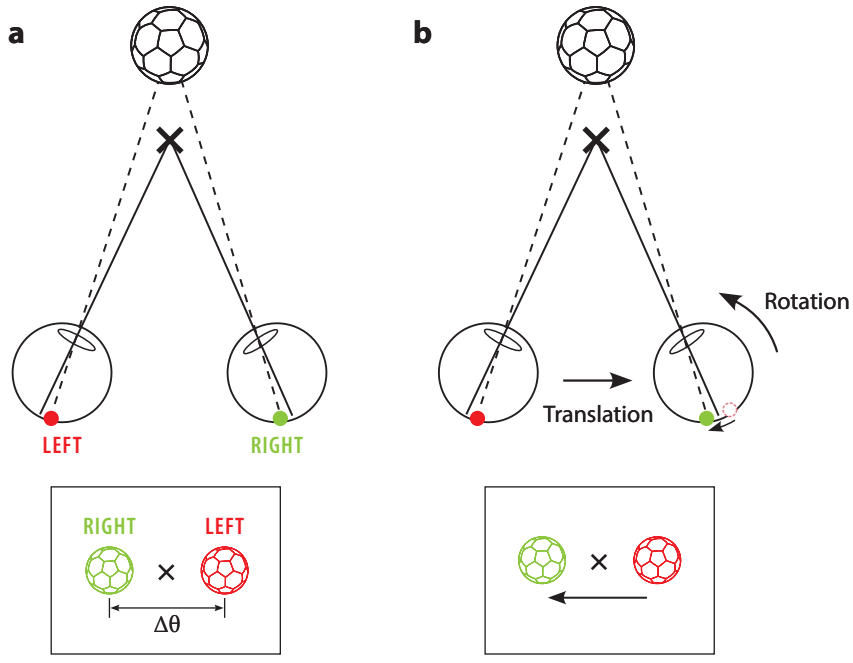

Figure 5

Similar geometries for binocular disparity and motion parallax cues. (*a*) Binocular disparity. When the eyes fixate on a stationary target (*black x*), the soccer ball projects onto different locations on the two retinae (*green* and *red dots*). The panel shows a top-down view of the geometry (*top*) and corresponding retinal images (*bottom*). (*b*) Motion parallax. When the observer translates laterally and rotates their eye to maintain visual fixation on the target (*black x*), the image projection of the soccer ball moves across the retina.

3.1. Geometry of Motion Parallax and Behavioral Evidence

The brain uses a variety of visual cues to infer the depth structure of a scene (Howard & Rogers 2008). These include pictorial cues such as shading, size, occlusion, and perspective, which we do not consider further here. Binocular disparity and motion parallax provide the most quantitative information about depth, as they both arise from viewing the scene from different vantage points. Indeed, the geometries underlying binocular disparity and motion parallax cues to depth are closely related. Binocular disparity refers to slight positional differences in images projected onto the two retinae and arises due to the horizontal separation between the eyes (**Figure 5*a***). In contrast, motion parallax refers to retinal image motion produced by combined translational and rotational movements of a single eye (**Figure 5*b***). When the eye translates laterally by an amount equivalent to the interpupillary distance and simultaneously counter-rotates to maintain fixation on a stationary target (as illustrated in **Figure 5*b***), the resulting image motion is equivalent to the binocular disparity of the same object viewed with both eyes (Bradshaw & Rogers 1996, Rogers & Graham 1982).

Since the work of von Helmholtz (1962), depth perception from motion parallax has been extensively studied. Using random-dot stimuli, researchers have carefully studied this monocular depth cue by isolating it from other depth cues, such as binocular disparity (Ono et al. 1986, Rogers 1993, Rogers & Graham 1979). In the viewing geometry of **Figure 5*b***, inferring depth from motion parallax requires knowledge of either head translation or eye rotation relative to the scene, if other depth cues are eliminated (Nawrot 2003). Studies have shown that smooth

pursuit eye movements play a key role in perceiving depth from motion parallax cues (Naji & Freeman 2004, Nawrot 2003). Later, Nawrot & Stroyan (2009) showed that the relative depth of a stationary object, d/f, can be approximated by the ratio between retinal velocity, v_{ret}, and eye velocity relative to the scene, v_{eye}:

$$\frac{d}{f} \approx \frac{v_{\text{ret}}}{v_{\text{eye}}}, \qquad 6.$$

where d is the distance between a stationary object and the fixation point and f is the viewing distance.

Many other species, such as pigeons (Frost 1978, Hataji et al. 2021, Xiao & Frost 2013), mice (Parker et al. 2022a), owls (van der Willigen et al. 2002), and some insects (Horridge 1986, Kral 2003, Kral & Poteser 1997, Preiss 1987), also use motion parallax cues to infer depth. Analogous parallax cues can also be found in other sensory modalities such as audition (Genzel et al. 2018, Yost 2018) and electric sensing (Hunke et al. 2021, Pedraja et al. 2018).

3.2. Neural Correlates in the Middle Temporal Area

In contrast to the vast literature on the neural coding of binocular disparity (for reviews, see Cumming & DeAngelis 2001, Parker 2007), relatively little was known about the neural coding of depth from motion parallax until more recently. Although some had speculated on a potential role of areas MT and MST in representing motion parallax (Cornilleau-Pérès & Gielen 1996), it was not until a decade later that Nadler et al. (2008) discovered neurons with tuning for depth defined by motion parallax in area MT. At this time, it remains unclear whether neurons with this property also exist in other visual areas. Consistent with theoretical and psychophysical results (Naji & Freeman 2004, Nawrot 2003, Nawrot & Stroyan 2009, Stroyan & Nawrot 2012), the extraretinal information that gives rise to selectivity for this depth cue in MT neurons involves signals related to smooth pursuit eye movements, rather than vestibular signals regarding head translation (Nadler et al. 2009). Importantly, this signal related to smooth pursuit can arise either from efference copy of motor commands (Nadler et al. 2009) or from large-field background motion that simulates eye rotation in this viewing geometry (Kim et al. 2015b). In addition, this novel selectivity for depth from motion parallax in MT neurons is integrated with selectivity for binocular disparity (Nadler et al. 2013) and has been functionally linked to depth discrimination behavior in macaques (Kim et al. 2015a). Thus, in addition to representing visual motion, area MT contains representations of depth based on both disparity and motion parallax.

How does information about smooth pursuit eye movements modulate MT neurons to generate selectivity for depth from motion parallax? An initial study suggested that depth tuning from motion parallax results from a gain modulation by eye movements (Kim et al. 2017). However, a more recent study shows that, for some MT neurons, depth selectivity appears to arise from a partial coordinate transformation from retinal to head-centered velocity coding (Xu & DeAngelis 2022). Consistent with recent findings regarding optic flow parsing (Peltier et al. 2024), these results indicate that area MT is involved in transforming motion signals into nonretinal reference frames.

3.3. Object Motion Biases Depth Perception

A crucial assumption exists in the motion-pursuit law, which describes how the depth of an object is computed from motion parallax (Nawrot & Stroyan 2009). This theory assumes that the object is stationary in the world, such that its retinal motion is solely due to self-motion (**Figure 6*a*,*d***). However, our visual environment often contains moving things (flying birds, moving cars, walking

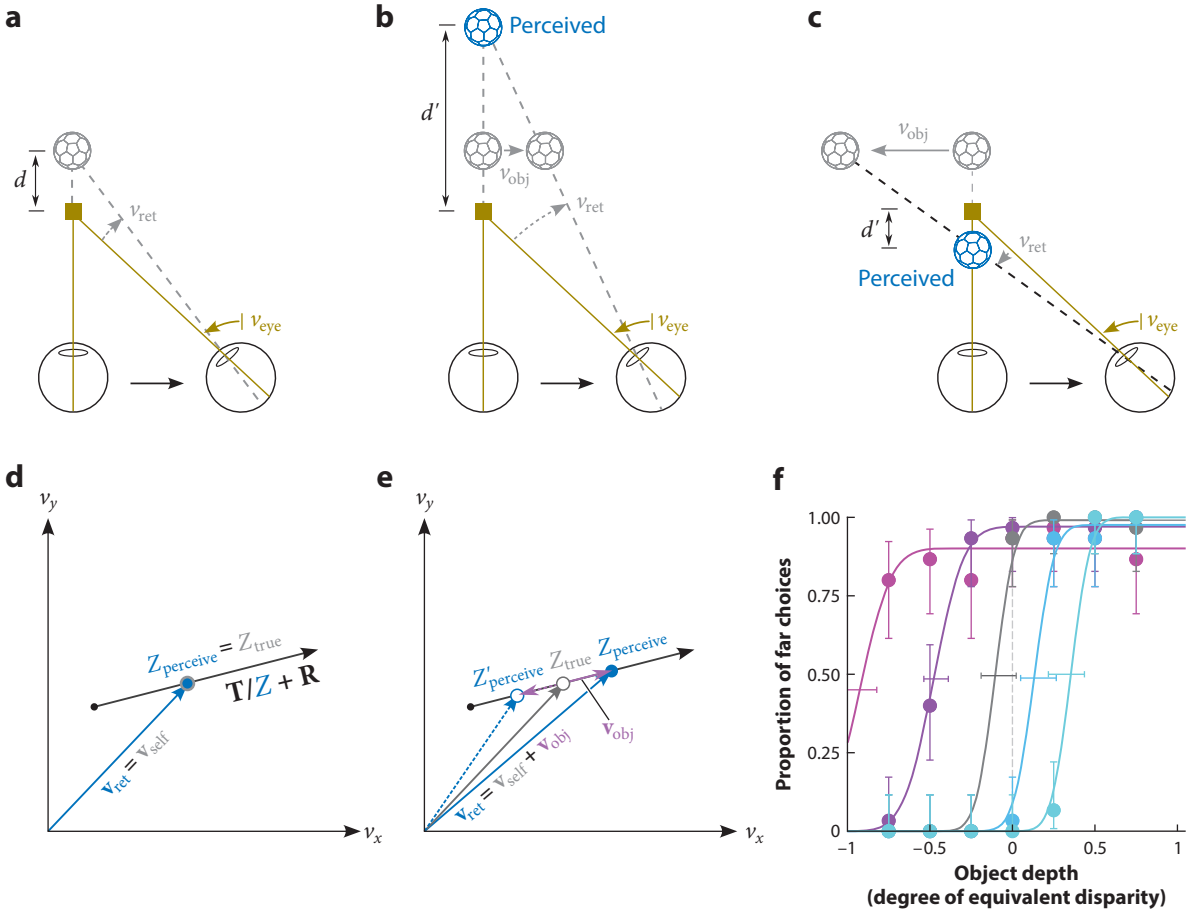

Figure 6

Object motion biases depth perception from motion parallax. (*a*) When an object (*soccer ball*) is stationary in the world, its depth can be accurately perceived from motion parallax cues. (*b*, *c*) When the object moves independently in the world, its motion contaminates retinal motion (v_{ret}, *dashed gray arrow*), causing a biased depth perception in the absence of other strong depth cues (*d′*, *blue soccer balls*). The direction of bias depends on the direction of object motion (v_{obj}, *solid gray arrows*). (*d*) Constraint ray analysis when the object is stationary. The object's depth is determined by the intersection (*blue dot*) between retinal velocity (*blue vector*) and the constraint ray (*black vector*). (*e*) Scene-relative object motion (*purple vectors*) shifts the intersection between retinal velocity and the constraint ray (*solid* and *open blue dots*), causing biased depth perception. (*f*) Psychometric functions for depth discrimination from an example subject. The gray curve shows the condition with a stationary object, the magenta curves show leftward object motion at different speeds, and the cyan curves show rightward object motion at different speeds. Error bars indicate 95% confidence intervals. Panel *f* adapted from French & DeAngelis (2022).

pedestrians, etc.). When the motion parallax cue to depth is contaminated by object motion, inferring depth becomes more challenging (French & DeAngelis 2020, 2022). Multiple interpretations are possible for the same viewing geometry and sensory observations (**Figure 6*b*,*c***). For example, the retinal motion of an object can be attributed to motion parallax, forming a perception of a stationary object at a particular depth (**Figure 6*b*,*c***, blue objects). It can also be explained as an independently moving object at a different depth (**Figure 6*b*,*c***, gray objects). A recent study showed that perception of an object's depth can be strongly biased by whether the object is moving relative to the scene (French & DeAngelis 2022) (**Figure 6*f***), demonstrating a coupling between the

perception of object motion and depth. This interaction can also be understood using the constraint ray (**Figure 6*e***): Object motion shifts the point at which the retinal velocity vector intersects the constraint ray. In the absence of other strong depth cues, this intersection point determines the perceived depth, such that object motion causes biased depth perception. By asking human subjects to simultaneously report the depth sign of an object (i.e., near versus far) and whether it is moving in the world, the perceptual tradeoffs between motion and depth can be modeled as a Bayesian causal inference problem (French 2021).

4. INFERENCE OF VIEWING GEOMETRY AND ITS ROLES IN MOTION AND DEPTH PERCEPTION

As discussed in Section 2.4, the effects of smooth pursuit eye movements on visual perception have been extensively studied in 2D scenarios (Freeman & Banks 1998; Freeman et al. 2000, 2010; Souman & Freeman 2008; Souman et al. 2006a; Spering & Montagnini 2011; Wertheim 1981, 1987, 1994). It is generally believed that the visual system subtracts a reference signal associated with the pursuit eye velocity from the retinal motion of an object to compute object motion in the world (Freeman & Banks 1998; Freeman et al. 2000, 2010; Souman & Freeman 2008; Souman et al. 2006a; Spering & Montagnini 2011; Wertheim 1981, 1987, 1994). A critical assumption of this type of model is that the visual consequence of smooth pursuit is a single optic flow vector that can be canceled by a reference signal (Freeman 2001, Wertheim 1994). While this assumption largely holds for pure eye rotations (**Figure 7*a***), the situation is more complicated when we consider a simple extension to eye rotation and translation against a 3D background (**Figure 7*d***).

First, consider the case in which a stationary observer makes a pure eye rotation to track a target against a 3D background (**Figure 7*a***, R viewing geometry). In this case, background image motion is a rotational flow field equivalent to a rotation of the scene around the eye (**Figure 7*a***). Under planar image projection, this rotational flow field is not purely laminar, but the motion vectors are independent of depth (in Equation 2, the depth-dependent term, $\mathbf{T}/Z$, is now zero, and $\mathbf{v}_{\text{self}} = \mathbf{R}$) (Kim et al. 2015b). Thus, the optic flow vector caused by smooth pursuit is well defined at any given point on the retina (**Figure 7*c***, v_{self}), and the classic model of subtracting a single reference vector is still applicable. For example, leftward smooth pursuit would induce a leftward bias in the perceived direction of a moving object relative to its retinal direction (**Figure 7*b***, orange curves), as observed empirically (Champion & Freeman 2010, Souman et al. 2005, Zivotofsky et al. 2005). Because the optic flow vectors do not vary as a function of depth in this viewing geometry, the observer would be unable to discriminate between near and far depths without additional depth cues (**Figure 7*i***, orange band).

This case of a pure eye rotation, however, is not typical of natural behavior. Rather, we often use smooth pursuit eye movements to transiently fixate a point in space (e.g., on the ground plane) while we translate relative to the scene (Lappi et al. 2020, Matthis et al. 2018, Tuhkanen et al. 2019). Even a microscopic head translation produces sufficient motion parallax cues for perceiving depth (Aytekin & Rucci 2012). This simple extension to a combination of eye rotation + translation (R+T) profoundly affects how the brain should compensate for the visual consequences of pursuit. In **Figure 7*d***, an observer translates laterally while counter-rotating their eye to maintain fixation on a world-fixed point (gold square). This R+T viewing geometry requires an identical smooth pursuit eye movement to the R geometry in **Figure 7*a***, but the visual consequences are strikingly different. In the R+T case, optic flow reflects a rotation of the scene around the point of fixation, such that near and far objects move in opposite directions and with speeds that increase with distance from fixation (**Figure 7*d***). In this viewing geometry, depth is related to the ratio of retinal and eye velocities (**Figure 7*e***), known as the motion-pursuit law (Nawrot & Stroyan 2009). Thus,

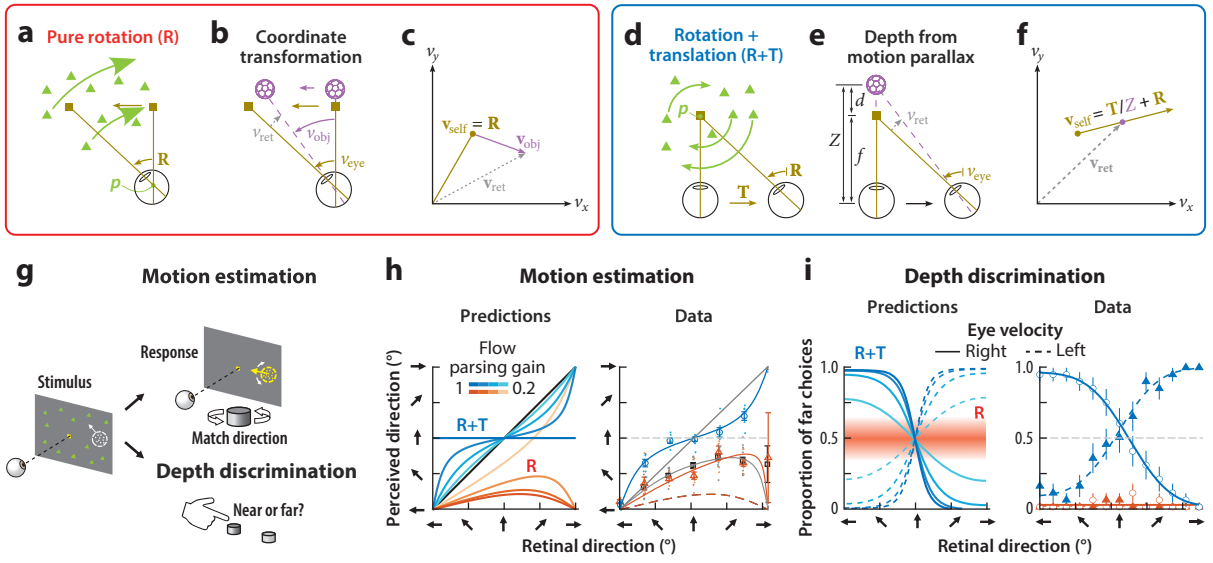

Figure 7

Perception of object motion and depth depends on the viewing geometry inferred from optic flow. (*a*) In the pure rotation (R) geometry, an observer rotates their eye to track a moving target (*gold square*). This generates optic flow (*green arrows*) that is depth invariant and can be described as a rotation of the scene around the center of the eye (pivot point, *p*). (*b*) In the R viewing geometry, it is natural for the brain to perform a coordinate transformation and compute object motion relative to the head or world. Here, the scene-relative velocity of an independently moving object (*soccer ball*, v_{obj}) can be inferred by adding eye velocity (v_{eye}, *gold arrow*) to retinal velocity (v_{ret}, *gray arrow*). (*c*) The constraint ray for the R geometry is reduced to a point (*gold dot*), and the object's motion is given by the violet vector. (*d*) In the rotation + translation (R+T) geometry, an observer translates to the right and counter-rotates their eye to maintain fixation at a stationary target (*gold square*). This yields optic flow that is highly depth dependent and can be described as a rotation of the scene around the fixation point, *p*. (*e*) In the R+T viewing geometry, it is natural for the brain to compute depth from motion parallax. By the motion-pursuit law, the relative depth of a stationary object (*soccer ball*), d/f, can be computed by taking the ratio between retinal and eye velocities. (*f*) In the R+T geometry, the absolute depth (*violet dot*), Z, is given by a particular point along the constraint ray defined by retinal motion. (*g*) Illustration of the motion estimation and depth discrimination tasks used in Xu et al. (2024). (*h*) Theoretical predictions for the motion estimation task (*left*) and data from an example subject (*right*). Note that data points have been slightly jittered horizontally for clarity. Different shadings represent varying amounts of self-motion accounted for by the observer. (*i*) Theoretical predictions (*left*) and data from an example subject (*right*) for the depth discrimination task. Error bars indicate the standard deviation of the mean. Panels *h* and *i* adapted from Xu et al. (2024) (CC BY-NC 4.0).

when the eye translates and counter-rotates horizontally, the horizontal component of an object's motion could be attributed to it lying at some particular depth, especially if other depth cues (e.g., size) are unreliable (**Figure 7*f***). Therefore, in this viewing geometry, it is natural for the brain to explain away at least some of the horizontal component of retinal motion as being due to depth, which should lead to a vertical bias in the perceived direction of an object (**Figure 7*h***, blue curves). Correspondingly, the sign of perceived depth (i.e., near versus far) should depend strongly on the direction of eye rotation in this viewing geometry (**Figure 7*i***, blue curves).

To summarize, even a simple extension toward a more natural viewing geometry (**Figure 7*d*,*e***) renders the classic scheme for pursuit compensation inadequate. This raises an important question: Does the brain automatically infer viewing geometry from optic flow and adjust its computations of motion and depth accordingly?

In an attempt to provide a unified account for motion and depth perception under different viewing geometries, Xu et al. (2024) characterized the relationship between viewing geometry and optic flow, demonstrating that humans automatically and flexibly perceive motion and depth

distinctly when different viewing geometries are simulated by optic flow. Specifically, they showed that the computations of motion and depth in these different viewing geometries can be unified by considering the rotation pivot of optic flow. In the R geometry, the optic flow rotates around the eye (**Figure 7*a***, point *p*), whereas in the R+T geometry, the optic flow rotates around the fixation point (**Figure 7*d***, point *p*). They derived a general relationship relating scene-relative object velocity, eye velocity, and retinal velocity:

$$v_{obj} = v_{ret} + (1 - (1 + d')\, p')\, v_{eye}. \qquad 7.$$

Here, $d' \triangleq d/f$, where f denotes the viewing distance (i.e., eye to fixation point), and d denotes depth relative to the fixation point (negative d indicates objects nearer than the fixation point). In addition, $p' \triangleq p/f$, where p denotes the point in space around which the scene rotates. For the R viewing geometry, $p' = 0$, and Equation 7 simplifies to $v_{obj} = v_{ret} + v_{eye}$ (Equation 5), which denotes a standard coordinate transformation from retinal to head coordinates (**Figure 7*b*,*c***). For the R+T viewing geometry, $p' = 1$, and Equation 7 simplifies to the standard motion-pursuit law (Nawrot & Stroyan 2009) when the target object is stationary in the world ($v_{obj} = 0$; **Figure 7*e***): $d' = d/f = v_{ret}/v_{eye}$.

By asking human subjects to provide an analog report of the motion direction of an object during the presence of different optic flow patterns (**Figure 7*g***, motion estimation), Xu et al. (2024) demonstrated that humans exhibit a horizontal bias in perceived direction when optic flow simulated the R viewing geometry and a vertical bias in perceived direction when optic flow simulated the R+T geometry (**Figure 7*h***, right), as predicted by the theoretical framework (**Figure 7*h***, left). Furthermore, results from a depth discrimination task (**Figure 7*g***) showed that reliable depth perception emerges in the R+T geometry but less so in the R geometry, with a strong relationship between the perceived depth sign (i.e., near versus far) and the direction of eye rotation in the R+T geometry (**Figure 7*i***, right) (Xu et al. 2024). This pattern of results is again well predicted by the theory (**Figure 7*i***, left).

These novel findings demonstrate that human observers automatically, and without any training or feedback, infer their viewing geometry from optic flow and alter their computations of object motion and depth. Understanding how these flexible computations are implemented neurally is of considerable interest (Xu 2023).

5. CONCLUSIONS

Existing visual psychophysical studies have primarily focused on relatively simple viewing geometries, such as a stationary observer making pursuit eye movements in studies of coordinate transformation (De Graaf & Wertheim 1988; Filehne 1922; Freeman & Banks 1998; Freeman et al. 2010; Souman et al. 2005, 2006a; Spering & Montagnini 2011; Wertheim 1981, 1987, 1994), an observer moving laterally and counter-rotating their eyes in studies of depth from motion parallax (Bradshaw & Rogers 1996; Gogel & Tietz 1973; Graham & Rogers 1982; Naji & Freeman 2004; Nawrot 2003; Nawrot & Stroyan 2009; Ono et al. 1986, 1988; Rogers & Graham 1982; Rogers & Rogers 1992; Stroyan & Nawrot 2012), or a forward-moving observer who makes no eye movements in studies of flow parsing (Foulkes et al. 2013; MacNeilage et al. 2012; Niehorster & Li 2017; Peltier et al. 2020; Warren & Rushton 2007, 2008, 2009a,b). We have reviewed our current understanding of how humans perceive object motion and depth under these constrained viewing geometries, along with what is known about some of the underlying neural mechanisms.

In natural behavior, however, our actions often include translations and rotations along multiple axes at the same time (Matthis et al. 2022), and the computations of object motion and depth are highly intertwined in such complex viewing geometries. Fortunately, optic flow generated by

self-motion is well characterized (Gibson 1950, Longuet-Higgins & Prazdny 1980), and our analysis of viewing geometry provides a framework that can begin to unify the interactions between object motion, depth, and self-motion under more natural conditions. Constraint ray analysis is a valuable tool that helps to reveal general principles that can be applied to self-motion in more naturalistic environments. The translation, **T**, and rotation, **R**, terms include movements in six degrees of freedom (Equation 2), and computations of object motion and depth may be performed for a variety of more complex types of self-motion. Our analysis thus provides a framework for quantitative predictions of motion and depth perception during self-motion. Future work focusing on movements that involve translations and rotations along multiple axes will be valuable for validating our framework. Virtual reality, along with eye tracking and omnidirectional treadmills, will be a powerful tool for such investigations.

Another step toward more naturalistic conditions involves characterizing the statistics of optic flow generated during natural behaviors (Matthis et al. 2018, 2022; Muller et al. 2023; Panfili et al. 2022). Recent studies have shown that retinal motion statistics during natural behavior reflect complex interactions between gaze location and locomotion, which are adaptively shaped by the environment (Matthis et al. 2018). It remains unknown whether perception of object motion and depth during natural behavior follows the same principles as identified in existing psychophysical studies.

SUMMARY POINTS

1. Visual motion provides powerful cues to infer object motion, depth, and self-motion, but the interactions between these three variables also present a major challenge for the analysis and interpretation of retinal image motion.
2. Substantial progress has been made in understanding how humans perceive motion or depth in simple viewing geometries, but less is known about the full three-way interactions between object motion, depth, and self-motion.
3. Neural mechanisms for computing object motion and depth have been elucidated in some specific self-motion conditions, but much remains to be learned about the underlying neural computations.
4. Quantitative analysis of how optic flow depends on viewing geometry makes specific predictions for how perception of object motion and depth is influenced by viewing geometry, and recent studies provide empirical support for these predictions.

FUTURE ISSUES

1. Further validation of our analysis framework requires testing the interactions between object motion, depth, and self-motion in more generalized viewing geometries.
2. Neural mechanisms underlying the full three-way interactions between object motion, depth, and self-motion remain largely unexplored.
3. Training animals to perform perceptual decision-making tasks that involve joint judgments of multiple variables will inform our understanding of the neural mechanisms of perceptual tradeoffs.

4. Investigation of the statistics of viewing geometries in natural behavior would facilitate the study of object motion and depth perception in natural settings.
5. Recording neural activity in freely moving animals may provide insights into the neural mechanisms by which self-motion is accounted for in all six degrees of freedom.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors were supported by National Institute of Neurological Disorders and Stroke U19 grant NS118246, as well as National Eye Institute R01 grants EY013644 and EY016178. We thank Akiyuki Anzai, Vitaly Lerner, and Yelin Dong for helpful comments on an earlier draft of the manuscript.

LITERATURE CITED

Acerbi L, Dokka K, Angelaki DE, Ma WJ. 2018. Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. *PLOS Comput. Biol.* 14:e1006110

Albright TD, Stoner GR. 1995. Visual motion perception. *PNAS* 92:2433–40

Andersen RA. 1989. Visual and eye movement functions of the posterior parietal cortex. *Annu. Rev. Neurosci.* 12:377–403

Andersen RA, Bradley DC. 1998. Perception of three-dimensional structure from motion. *Trends Cogn. Sci.* 2:222–28

Andersen RA, Mountcastle VB. 1983. The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J. Neurosci.* 3:532–48

Andersen RA, Zipser D. 1988. The role of the posterior parietal cortex in coordinate transformations for visual–motor integration. *Can. J. Physiol. Pharm.* 66:488–501

Aubert H. 1887. Die Bewegungsempfindung. *Pflügers Arch.* 40:459–80

Aytekin M, Rucci M. 2012. Motion parallax from microscopic head movements during visual fixation. *Vis. Res.* 70:7–17

Barron JL, Fleet DJ, Beauchemin SS. 1994. Performance of optical flow techniques. *Int. J. Comp. Vis.* 12:43–77

Born RT, Bradley DC. 2005. Structure and function of visual area MT. *Annu. Rev. Neurosci.* 28:157–89

Bradshaw MF, Rogers BJ. 1996. The interaction of binocular disparity and motion parallax in the computation of depth. *Vis. Res.* 36:3457–68

Britten KH. 2008. Mechanisms of self-motion perception. *Annu. Rev. Neurosci.* 31:389–410

Brostek L, Büttner U, Mustari MJ, Glasauer S. 2015. Eye velocity gain fields in MSTd during optokinetic stimulation. *Cereb. Cortex* 25:2181–90

Champion RA, Freeman TCA. 2010. Discrimination contours for the perception of head-centered velocity. *J. Vis.* 10(6):14

Chukoskie L, Movshon JA. 2009. Modulation of visual signals in macaque MT and MST neurons during pursuit eye movement. *J. Neurophysiol.* 102:3225–33

Cornilleau-Pérès V, Gielen CCAM. 1996. Interactions between self-motion and depth perception in the processing of optic flow. *Trends Neurosci.* 19:196–202

Cumming BG, DeAngelis GC. 2001. The physiology of stereopsis. *Annu. Rev. Neurosci.* 24:203–38

De Graaf B, Wertheim AH. 1988. The perception of object motion during smooth pursuit eye movements: adjacency is not a factor contributing to the Filehne illusion. *Vis. Res.* 28:497–502

DeAngelis GC, Newsome WT. 1999. Organization of disparity-selective neurons in macaque area MT. *J. Neurosci.* 19:1398–415
DeAngelis GC, Uka T. 2003. Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J. Neurophysiol.* 89:1094–111
DiRisio GF, Ra Y, Qiu Y, Anzai A, DeAngelis GC. 2023. Neurons in primate area MSTd signal eye movement direction inferred from dynamic perspective cues in optic flow. *J. Neurosci.* 43:1888–904
Dokka K, DeAngelis GC, Angelaki DE. 2015a. Multisensory integration of visual and vestibular signals improves heading discrimination in the presence of a moving object. *J. Neurosci.* 35:13599–607
Dokka K, MacNeilage PR, DeAngelis GC, Angelaki DE. 2015b. Multisensory self-motion compensation during object trajectory judgments. *Cereb. Cortex* 25:619–30
Dokka K, Park H, Jansen M, DeAngelis GC, Angelaki DE. 2019. Causal inference accounts for heading perception in the presence of object motion. *PNAS* 116:9060–65
Fajen BR, Kim N-G. 2002. Perceiving curvilinear heading in the presence of moving objects. *J. Exp. Psychol. Hum. Percept. Perform.* 28:1100–19
Festinger L, Sedgwick HA, Holtzman JD. 1976. Visual perception during smooth pursuit eye movements. *Vis. Res.* 16:1377–86
Fetsch CR, Turner AH, DeAngelis GC, Angelaki DE. 2009. Dynamic reweighting of visual and vestibular cues during self-motion perception. *J. Neurosci.* 29:15601–12
Filehne W. 1922. Über das optische Wahrnehmen von Bewegungen. *Z. Sinnesphysiol.* 53:134–45
Foulkes AJ, Rushton SK, Warren PA. 2013. Flow parsing and heading perception show similar dependence on quality and quantity of optic flow. *Front. Behav. Neurosci.* 7:49
Freeman TCA. 2001. Transducer models of head-centred motion perception. *Vis. Res.* 41:2741–55
Freeman TCA, Banks MS. 1998. Perceived head-centric speed is affected by both extra-retinal and retinal errors. *Vis. Res.* 38:941–45
Freeman TCA, Banks MS, Crowell JA. 2000. Extraretinal and retinal amplitude and phase errors during Filehne illusion and path perception. *Percept. Psychophys.* 62:900–9
Freeman TCA, Champion RA, Warren PA. 2010. A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Curr. Biol.* 20:757–62
French RL. 2021. *Perceptual processes underlying depth judgements of moving objects during self-motion*. PhD Diss., University of Rochester
French RL, DeAngelis GC. 2020. Multisensory neural processing: from cue integration to causal inference. *Curr. Opin. Physiol.* 16:8–13
French RL, DeAngelis GC. 2022. Scene-relative object motion biases depth percepts. *Sci. Rep.* 12:18480
Frost BJ. 1978. The optokinetic basis of head-bobbing in the pigeon. *J. Exp. Biol.* 74:187–95
Genzel D, Schutte M, Brimijoin WO, MacNeilage PR, Wiegrebe L. 2018. Psychophysical evidence for auditory motion parallax. *PNAS* 115:4264–69
Gibson EJ, Gibson JJ, Smith OW, Flock H. 1959. Motion parallax as a determinant of perceived depth. *J. Exp. Psychol.* 58:40–51
Gibson JJ. 1950. *The Perception of the Visual World*. Houghton Mifflin
Gogel WC, Tietz JD. 1973. Absolute motion parallax and the specific distance tendency. *Percept. Psychophys.* 13:284–92
Gogel WC, Tietz JD. 1979. A comparison of oculomotor and motion parallax cues of egocentric distance. *Vis. Res.* 19:1161–70
Graham M, Rogers B. 1982. Simultaneous and successive contrast effects in the perception of depth from motion-parallax and stereoscopic information. *Perception* 11:247–62
Groh JM, Born RT, Newsome WT. 1997. How is a sensory map read out? Effects of microstimulation in visual area MT on saccades and smooth pursuit eye movements. *J. Neurosci.* 17:4312–30
Gu Y, Angelaki DE, DeAngelis GC. 2008. Neural correlates of multisensory cue integration in macaque MSTd. *Nat. Neurosci.* 11:1201–10
Hataji Y, Kuroshima H, Fujita K. 2021. Motion parallax via head movements modulates visuo-motor control in pigeons. *J. Exp. Biol.* 224:jeb236547
Horn BKP, Schunck BG. 1981. Determining optical flow. *Artif. Intell.* 17:185–203

Horridge GA. 1986. A theory of insect vision: velocity parallax. *Proc. R. Soc. B* 229:13–27
Howard IP, Rogers BJ. 2008. *Seeing in Depth*. Oxford University Press
Hunke K, Engelmann J, Meyer HG, Schneider A. 2021. Motion parallax for object localization in electric fields. *Bioinspir. Biomim.* 17:016003
Ilg UJ, Schumann S, Thier P. 2004. Posterior parietal cortex neurons encode target motion in world-centered coordinates. *Neuron* 43:145–51
Inaba N, Miura K, Kawano K. 2011. Direction and speed tuning to visual motion in cortical areas MT and MSTd during smooth pursuit eye movements. *J. Neurophysiol.* 105:1531–45
Inaba N, Shinomoto S, Yamane S, Takemura A, Kawano K. 2007. MST neurons code for visual motion in space independent of pursuit eye movements. *J. Neurophysiol.* 97:3473–83
Jain RC. 1984. Segmentation of frame sequences obtained by a moving observer. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-6:624–29
Johansson G. 1975. Visual motion perception. *Sci. Am.* 232:76–89
Jörges B, Harris LR. 2022. Object speed perception during lateral visual self-motion. *Attent. Percept. Psychophys.* 84:25–46
Kim HR, Angelaki DE, DeAngelis GC. 2015a. A functional link between MT neurons and depth perception based on motion parallax. *J. Neurosci.* 35:2766–77
Kim HR, Angelaki DE, DeAngelis GC. 2015b. A novel role for visual perspective cues in the neural computation of depth. *Nat. Neurosci.* 18:129–37
Kim HR, Angelaki DE, DeAngelis GC. 2016. The neural basis of depth perception from motion parallax. *Philos. Trans. R. Soc. B* 371:20150256. Correction. 2016. *Philos. Trans. R. Soc. B* 371:20160395
Kim HR, Angelaki DE, DeAngelis GC. 2017. Gain modulation as a mechanism for coding depth from motion parallax in macaque area MT. *J. Neurosci.* 37:8180–97
Kim HR, Angelaki DE, DeAngelis GC. 2022. A neural mechanism for detecting object motion during self-motion. *eLife* 11:e74971
Koenderink JJ. 1986. Optic flow. *Vis. Res.* 26:161–79
Koenderink JJ, van Doorn AJ. 1991. Affine structure from motion. *J. Opt. Soc. Am. A* 8:377–85
Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. 2007. Causal inference in multisensory perception. *PLOS ONE* 2:e943
Kral K. 2003. Behavioural–analytical studies of the role of head movements in depth perception in insects, birds and mammals. *Behav. Process.* 64:1–12
Kral K, Poteser M. 1997. Motion parallax as a source of distance information in locusts and mantids. *J. Insect Behav.* 10:145–63
Lappe M, Bremmer F, van den Berg AV. 1999. Perception of self-motion from visual flow. *Trends Cogn. Sci.* 3:329–36
Lappi O, Pekkanen J, Rinkkala P, Tuhkanen S, Tuononen A, Virtanen J-P. 2020. Humans use optokinetic eye movements to track waypoints for steering. *Sci. Rep.* 10:4175
Layton OW, Fajen BR. 2016a. Sources of bias in the perception of heading in the presence of moving objects: object-based and border-based discrepancies. *J. Vis.* 16(1):9
Layton OW, Fajen BR. 2016b. The temporal dynamics of heading perception in the presence of moving objects. *J. Neurophysiol.* 115:286–300
Lisberger SG. 2015. Visual guidance of smooth pursuit eye movements. *Annu. Rev. Vis. Sci.* 1:447–68
Lisberger SG, Morris EJ, Tychsen L. 1987. Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annu. Rev. Neurosci.* 10:97–129
Longuet-Higgins HC, Prazdny K. 1980. The interpretation of a moving retinal image. *Proc. R. Soc. B* 208:385–97
Lu Z-L, Sperling G. 1995. The functional architecture of human visual motion perception. *Vis. Res.* 35:2697–722
Lucas BD, Kanade T. 1981. An iterative image registration technique with an application to stereo vision. Paper presented at 7th International Joint Conference on Artificial Intelligence, Aug. 1981
Lutwak H, Rokers B, Simoncelli EP. 2025. Detection of moving objects using self-motion constraints on optic flow. Preprint, arXiv:2505.06686v2 [q-bio.NC]

Mack A, Herman E. 1973. Position constancy during pursuit eye movement: an investigation of the Filehne illusion. *Q. J. Exp. Psychol.* 25:71–84
MacNeilage PR, Zhang Z, DeAngelis GC, Angelaki DE. 2012. Vestibular facilitation of optic flow parsing. *PLOS ONE* 7:e40264
Matin E. 1974. Saccadic suppression: a review and an analysis. *Psychol. Bull.* 81:899–917
Matthis JS, Muller KS, Bonnen KL, Hayhoe MM. 2022. Retinal optic flow during natural locomotion. *PLOS Comput. Biol.* 18:e1009575
Matthis JS, Yates JL, Hayhoe MM. 2018. Gaze and the control of foot placement when walking in natural terrain. *Curr. Biol.* 28:1224–33.e5
Maunsell JH, Van Essen DC. 1983. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J. Neurophysiol.* 49:1148–67
Mayer KM, Riddell H, Lappe M. 2021. Flow parsing and biological motion. *Atten. Percept. Psychophys.* 83:1752–65
Morgan ML, DeAngelis GC, Angelaki DE. 2008. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* 59:662–73
Morris AP, Krekelberg B. 2019. A stable visual world in primate primary visual cortex. *Curr. Biol.* 29:1471–80.e6
Muller KS, Matthis J, Bonnen K, Cormack LK, Huk AC, Hayhoe M. 2023. Retinal motion statistics during natural locomotion. *eLife* 12:e82410
Nadler JW, Angelaki DE, DeAngelis GC. 2008. A neural representation of depth from motion parallax in macaque visual cortex. *Nature* 452:642–45
Nadler JW, Barbash D, Kim HR, Shimpi S, Angelaki DE, DeAngelis GC. 2013. Joint representation of depth from motion parallax and binocular disparity cues in macaque area MT. *J. Neurosci.* 33:14061–74
Nadler JW, Nawrot M, Angelaki DE, DeAngelis GC. 2009. MT neurons combine visual motion with a smooth eye movement signal to code depth-sign from motion parallax. *Neuron* 63:523–32
Naji JJ, Freeman TCA. 2004. Perceiving depth order during pursuit eye movement. *Vis. Res.* 44:3025–34
Nawrot M. 2003. Eye movements provide the extra-retinal signal required for the perception of depth from motion parallax. *Vis. Res.* 43:1553–62
Nawrot M, Blake R. 1989. Neural integration of information specifying structure from stereopsis and motion. *Science* 244:716–18
Nawrot M, Joyce L. 2006. The pursuit theory of motion parallax. *Vis. Res.* 46:4709–25
Nawrot M, Stroyan K. 2009. The motion/pursuit law for visual depth perception from motion parallax. *Vis. Res.* 49:1969–78
Nelson RC. 1991. Qualitative detection of motion by a moving observer. *Int. J. Comput. Vis.* 7:33–46
Niehorster DC, Li L. 2017. Accuracy and tuning of flow parsing for visual perception of object motion during self-motion. *i-Perception* 8:2041669517708206
Nishida S, Kawabe T, Sawayama M, Fukiage T. 2018. Motion perception: from detection to interpretation. *Annu. Rev. Vis. Sci.* 4:501–23
Ono H, Rogers BJ, Ohmi M, Ono ME. 1988. Dynamic occlusion and motion parallax in depth perception. *Perception* 17:255–66
Ono ME, Rivest J, Ono H. 1986. Depth perception as a function of motion parallax and absolute-distance information. *J. Exp. Psychol. Hum. Percept. Perform.* 12:331–37
Panfili D, Muller K, Hayhoe M. 2022. Flow segmentation during locomotion. *J. Vis.* 22(14):4040
Park WJ, Tadin D. 2018. Motion perception. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, Vol. 2, ed. JT Serences, JT Wixted. John Wiley & Sons
Parker AJ. 2007. Binocular depth perception and the cerebral cortex. *Nat. Rev. Neurosci.* 8:379–91
Parker PRL, Abe ETT, Beatie NT, Leonard ESP, Martins DM, et al. 2022a. Distance estimation from monocular cues in an ethological visuomotor task. *eLife* 11:e74708
Parker PRL, Abe ETT, Leonard ESP, Martins DM, Niell CM. 2022b. Joint coding of visual input and eye/head position in V1 of freely moving mice. *Neuron* 110:3897–906.e5
Pedraja F, Hofmann V, Lucas KM, Young C, Engelmann J, Lewis JE. 2018. Motion parallax in electric sensing. *PNAS* 115:573–77

Peltier NE, Angelaki DE, DeAngelis GC. 2020. Optic flow parsing in the macaque monkey. *J. Vis.* 20(10):8
Peltier NE, Anzai A, Moreno-Bote R, DeAngelis GC. 2024. A neural mechanism for optic flow parsing in macaque visual cortex. *Curr. Biol.* 34:4983–97.e9
Pouget A, Sejnowski TJ. 1997. Spatial transformations in the parietal cortex using basis functions. *J. Cogn. Neurosci.* 9:222–37
Preiss R. 1987. Motion parallax and figural properties of depth control flight speed in an insect. *Biol. Cybern.* 57:1–9
Rogers B. 2009. Motion parallax as an independent cue for depth perception: a retrospective. *Perception* 38:907–11
Rogers B, Graham M. 1979. Motion parallax as an independent cue for depth perception. *Perception* 8:125–34
Rogers B, Graham M. 1982. Similarities between motion parallax and stereopsis in human depth perception. *Vis. Res.* 22:261–70
Rogers BJ. 1993. Motion parallax and other dynamic cues for depth in humans. *Rev. Oculomot. Res.* 5:119–37
Rogers S, Rogers BJ. 1992. Visual and nonvisual information disambiguate surfaces specified by motion parallax. *Percept. Psychophys.* 52:446–52
Royden CS, Connors EM. 2010. The detection of moving objects by moving observers. *Vis. Res.* 50:1014–24
Rushton SK, Bradshaw MF, Warren PA. 2007. The pop out of scene-relative object movement against retinal motion due to self-movement. *Cognition* 105:237–45
Rushton SK, Niehorster DC, Warren PA, Li L. 2018. The primary role of flow processing in the identification of scene-relative object movement. *J. Neurosci.* 38:1737–43
Salinas E, Abbott LF. 1995. Transfer of coded information from sensory to motor networks. *J. Neurosci.* 15:6461–74
Sasaki R, Anzai A, Angelaki DE, DeAngelis GC. 2020. Flexible coding of object motion in multiple reference frames by parietal cortex neurons. *Nat. Neurosci.* 23:1004–15
Shadlen MN, Newsome WT. 1996. Motion perception: seeing and deciding. *PNAS* 93:628–33
Shams L, Beierholm UR. 2010. Causal inference in perception. *Trends Cogn. Sci.* 14:425–32
Smith OW, Smith PC. 1963. On motion parallax and perceived depth. *J. Exp. Psychol.* 65:107–8
Souman JL, Freeman TCA. 2008. Motion perception during sinusoidal smooth pursuit eye movements: signal latencies and non-linearities. *J. Vis.* 8(14):10
Souman JL, Hooge ITC, Wertheim AH. 2005. Vertical object motion during horizontal ocular pursuit: compensation for eye movements increases with presentation duration. *Vis. Res.* 45:845–53
Souman JL, Hooge ITC, Wertheim AH. 2006a. Frame of reference transformations in motion perception during smooth pursuit eye movements. *J. Comput. Neurosci.* 20:61–76
Souman JL, Hooge ITC, Wertheim AH. 2006b. Localization and motion perception during smooth pursuit eye movements. *Exp. Brain Res.* 171:448–58
Spering M, Montagnini A. 2011. Do we track what we see? Common versus independent processing for motion perception and smooth pursuit eye movements: a review. *Vis. Res.* 51:836–52
Stocker AA, Simoncelli EP. 2006. Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* 9:578–85
Stone LS, Beutter BR, Lorenceau J. 2000. Visual motion integration for perception and pursuit. *Perception* 29:771–87
Stroyan K, Nawrot M. 2012. Visual depth from motion parallax and eye pursuit. *J. Math. Biol.* 64:1157–88
Thompson WB, Pong T-C. 1990. Detecting moving objects. *Int. J. Comput. Vis.* 4:39–57
Tuhkanen S, Pekkanen J, Rinkkala P, Mole C, Wilkie RM, Lappi O. 2019. Humans use predictive gaze strategies to target waypoints for steering. *Sci. Rep.* 9:8344
Uka T, DeAngelis GC. 2003. Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. *J. Neurosci.* 23:3515–30
Ullman S. 1979. The interpretation of structure from motion. *Proc. R. Soc. B* 203:405–26
van der Willigen RF, Frost BJ, Wagner H. 2002. Depth generalization from stereo to motion parallax in the owl. *J. Comp. Physiol. A* 187:997–1007
von Fleischl E. 1882. *Physiologisch-optische Notizen*. k. k. Hof- und Staatsdruckerei
von Helmholtz H. 1962. *Helmholtz's Treatise on Physiological Optics.* Dover Publications

Warren PA, Rushton SK. 2007. Perception of object trajectory: parsing retinal motion into self and object movement components. *J. Vis.* 7(11):2

Warren PA, Rushton SK. 2008. Evidence for flow-parsing in radial flow displays. *Vis. Res.* 48:655–63

Warren PA, Rushton SK. 2009a. Optic flow processing for the assessment of object movement during ego movement. *Curr. Biol.* 19:1555–60

Warren PA, Rushton SK. 2009b. Perception of scene-relative object movement: optic flow parsing and the contribution of monocular depth cues. *Vis. Res.* 49:1406–19

Warren WH Jr., Hannon DJ. 1988. Direction of self-motion is perceived from optical flow. *Nature* 336:162–63

Wertheim AH. 1981. On the relativity of perceived motion. *Acta Psychol.* 48:97–110

Wertheim AH. 1987. Retinal and extraretinal information in movement perception: how to invert the Filehne illusion. *Perception* 16:299–308

Wertheim AH. 1994. Motion perception during self-motion: the direct versus inferential controversy revisited. *Behav. Brain Sci.* 17:293–311

Xiao Q, Frost BJ. 2013. Motion parallax processing in pigeon (*Columba livia*) pretectal neurons. *Eur. J. Neurosci.* 37:1103–11

Xu Z. 2023. *Eyes in motion: computational principles and neural basis of motion and depth perception by a moving observer*. PhD Thesis, University of Rochester

Xu Z-X, DeAngelis GC. 2022. Neural mechanism for coding depth from motion parallax in area MT: gain modulation or tuning shifts? *J. Neurosci.* 42:1235–53

Xu Z-X, Pang J, Anzai A, DeAngelis GC. 2024. Flexible computation of object motion and depth based on viewing geometry inferred from optic flow. Preprint, bioRxiv. **https://www.biorxiv.org/content/10.1101/2024.10.29.620928v2.abstract**

Yarbus AL. 2013. *Eye Movements and Vision*. Springer

Yost WA. 2018. Auditory motion parallax. *PNAS* 115:3998–4000

Zeki S. 1980. The response properties of cells in the middle temporal area (area MT) of owl monkey visual cortex. *Proc. R. Soc. B* 207:239–48

Zivotofsky AZ, Goldberg ME, Powell KD. 2005. Rhesus monkeys behave as if they perceive the Duncker Illusion. *J. Cogn. Neurosci.* 17:1011–17