

A motion pooling model of visually guided navigation explains human behavior in the presence of independently moving objects

Oliver W. Layton

Center for Computational Neuroscience and Neural Technology, Boston University, Boston, MA, USA, & Program in Cognitive and Neural Systems, Boston University, Boston, MA, USA



Ennio Mingolla

Center for Computational Neuroscience and Neural Technology, Boston University, Boston, MA, USA, & Department of Psychology, Boston University, Boston, MA, USA



N. Andrew Browning

Center for Computational Neuroscience and Neural Technology, Boston University, Boston, MA, USA



Humans accurately judge their direction of heading when translating in a rigid environment, unless independently moving objects (IMOs) cross the observer's focus of expansion (FoE). Studies show that an IMO on a laterally moving path that maintains a fixed distance with respect to the observer (non-approaching; C. S. Royden & E. C. Hildreth, 1996) biases human heading estimates differently from an IMO on a lateral path that gets closer to the observer (approaching; W. H. Warren & J. A. Saunders, 1995). C. S. Royden (2002) argued that differential motion operators in primate brain area MT explained both data sets, concluding that differential motion was critical to human heading estimation. However, neurophysiological studies show that motion pooling cells, but not differential motion cells, in MT project to heading-sensitive cells in MST (V. K. Berezovskii & R. T. Born, 2000). It is difficult to reconcile differential motion heading models with these neurophysiological data. We generate motion sequences that mimic those viewed by human subjects. Model MT pools over V1; units in model MST perform distance-weighted template matching and compete in a recurrent heading representation layer. Our model produces heading biases of the same direction and magnitude as humans through a peak shift in model MSTd without using differential motion operators, maintaining consistency with known primate neurophysiology.

Keywords: optic flow, independently moving object, heading, MT, MST, navigation, neural model

Citation: Layton, O. W., Mingolla, E., & Browning, N. A. (2012). A motion pooling model of visually guided navigation explains human behavior in the presence of independently moving objects. *Journal of Vision*, 12(1):20, 1–19, <http://www.journalofvision.org/content/12/1/20>, doi:10.1167/12.1.20.

Introduction

Navigation is an important activity of many species. Understanding how humans and other animals adeptly move around in their environments has stimulated much research (Warren, 2009; Warren, Kay, Zosh, Duchon, & Sahuc, 2001). Humans can accurately estimate their direction of heading in stationary environments to within 1–2° using visual information alone (Hatsopoulos & Warren, 1991; Royden, Crowell, & Banks, 1994; Warren, Blackwell, Kurtz, Hatsopoulos, & Kalish, 1991; Warren & Hannon, 1990; Warren & Kurtz, 1992; Warren, Morris, & Kalish, 1988). Animals with eyes sample structured distributions of light over time (called optic flow) to obtain information about heading (Gibson, 1979). During forward locomotion in stationary environments, and in the absence of sources of rotation due to eye, head, or body

movements, the observer experiences purely translational optic flow. Under such circumstances, the focus of expansion (FoE), the singularity in the radially expanding optic flow field, uniquely specifies the heading and direction of the linear path traversed by the observer. Although the FoE is only “visible” during forward locomotion, the optic flow field also contains a locus of inflow or focus of contraction (FoC; i.e., where one came from), which is visible during backward locomotion (Gibson, 1979). Animals with wide fields of view may be able to view both FoE and FoC simultaneously. Humans and other animals must also navigate in dynamic environments, often with independently moving objects (IMOs). In the presence of large IMOs, heading estimation performance remains good, but humans make systematic errors in the judged heading direction and magnitude under some circumstances (Fajen & Kim, 2002). We present a computational neural model that explains the patterns of

errors humans make in rigid environments to clarify the functional characteristics of brain circuits involved in the estimation of heading. The model performs as well in replicating the human heading estimation data from Royden and Hildreth (1996) and Warren and Saunders (1995) as the biologically inspired model of Royden (2002) but is more representative of known neurophysiology.

Theories of heading estimation

There are two main classes of human heading estimation models: *motion pooling models* and *differential motion models*. The focus of these heading models is typically to explain how the FoE can be used as an indicator of heading in a general case when rotations may be present in the visual field—as is often the case with animals that can move their eyes. Motion pooling models estimate the observer's heading by integrating motion signals over large portions of the visual field. The assumption of such models is that either the effects of rotation in the flow field on heading estimates is unimportant or that rotation has been previously removed from the optic flow field, for example, by using vestibular oculomotor signals (Pack, Grossberg, & Mingolla, 2001). Differential motion models exploit the separability of the translation and rotation components of optic flow as demonstrated by Longuet-Higgins and Prazdny's (1980) retinal flow equations to remove the effects of rotation from the motion field:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix} \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} + \frac{1}{f} \begin{pmatrix} xy & -(f^2 + x^2) & fy \\ (f^2 + y^2) & -xy & -fx \end{pmatrix} \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix}. \quad (1)$$

Equation 1 describes the instantaneous (i.e., first-order) optic flow for points in the world (X, Y, Z) projected onto a 2D “retinal plane” with coordinates (x, y), assuming a planar camera model (Longuet-Higgins & Prazdny, 1980). Each point in world coordinates has a depth Z , f denotes the camera focal length, and \vec{t} and \vec{r} represent the translational and rotational velocities of the observer, respectively. Suppose one samples two closely spaced motion vectors from the global array in a rigid environment at a location at which the depth changes (a “depth discontinuity”). Assuming the vectors are sufficiently close in space, they should have similar translational and rotational components but different depths. Since in **Equation 1** the rotational term does not depend on depth Z , but the translational term does, subtracting the two vectors yields the difference in translational components,

which is proportional to the size of the depth discontinuity. Supposing one samples and performs vector subtraction at several locations, one can analyze the difference vectors to recover the FoE due to observer translation. Differential motion models perform vector subtraction at depth discontinuities resulting in difference vectors that have the same angle as the original vectors, provided that the scene is rigid and the only motion is due to the observer translation (Perrone & Krauzlis, 2008). All that remains is to triangulate the “difference vectors” to obtain the estimated heading. While the angle of the difference vectors remains the same, the sign may be switched. Hence, one must have some knowledge about the direction of observer translation to differentiate the FoE from the FoC. The differential motion approach critically relies on the separability of translational and rotational components in **Equation 1** and on the existence of significant depth discontinuities (not gradients) in the environment. For a comprehensive review of methods used to analyze **Equation 1** and their performance, see Raudies and Neumann (accepted for publication).

IMOs may get closer to the observer, either due to observer or object motion (approaching IMO), or maintain a fixed distance from the observer irrespective of the observer motion (non-approaching IMO). In this article, we define approaching and non-approaching IMOs as the stimuli described by Royden and Hildreth (1996) and Warren and Saunders (1995), respectively. An approaching object is the natural case; when the observer moves toward the object, the distance between the observer and the object gets shorter. This may occur when driving a vehicle on a straight course and a truck enters the vehicle's future path from a perpendicular side street. A non-approaching object is less likely to occur in a natural setting; the motion of the observer in depth is the same as that of the IMO in that direction, as shown in **Figure 1a**. This may occur when driving a vehicle on a straight course and a car traveling at the same speed changes lanes from an adjacent lane, quickly passing in front of the observer. In both cases, the optic flow describing the object motion has an FoE that defines where the object is coming from. During observer translation in the presence of an IMO, the FoE due to observer motion and the FoE due to object motion are both present in the motion field. The goal of heading estimation is to extract the FoE due to observer motion. Humans do this very well unless the IMO crosses the observer FoE and thereby reduces its visibility. Human psychophysics data provide some design constraints that we can use to determine the nature and form of the neural circuits of the primate brain that give rise to heading perception.

Neurophysiological background

Neurons in primate medial temporal (MT) area are functionally tuned to properties including retinal position,

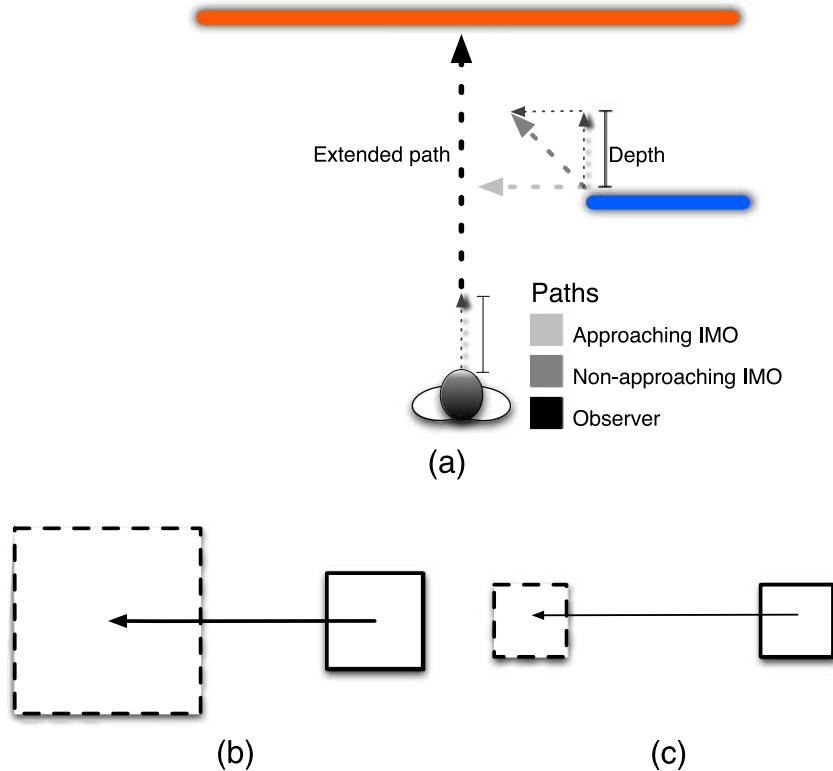


Figure 1. Schematic of approaching and non-approaching IMO conditions. (a) An observer translating toward a dot plane and an approaching IMO (light gray arrow) or a non-approaching IMO (medium gray arrow), which has a depth component of motion that equals that of the observer. (b) An approaching IMO. The solid and dashed outlines indicate its starting and ending appearances to the observer, respectively. (c) A non-approaching IMO. The solid and dashed outlines indicate its starting and ending appearances to the observer, respectively.

direction of motion, and speed (Born & Bradley, 2005). Like primary visual cortex (V1), each hemisphere features a retinotopic organization and nearly complete representation of the contralateral visual field (Gattass & Gross, 1981). V1 provides more inputs than any other area to MT (Nassi & Callaway, 2006; Sincich & Horton, 2005). Neurons in MT tend to possess receptive field sizes up to ten times larger than V1, typically 0.2°–1.2° (Born & Bradley, 2005; Zhou, Friedman, & von der Heydt, 2000). Early researchers suspected that MT participates in longer range motion integration than V1 does; however, recent work shows that cells in MT integrate motion over a shorter range than previously thought (Born & Bradley, 2005). Churchland, Priebe, Lisberger, Priebe, and Lis (2005) found that of a sample of 100 MT neurons, average second-order maps yielded spatial two-dot flash integrations of 0.62° and 0.73° for 16°/s and 32°/s flashes, respectively. These averages fell within 0.05° from those of sampled V1 neurons, showing sampled cells in MT appear to integrate motion over similar regions of space, despite having larger receptive fields.

Primate area MT contains at least two distinct populations of cells—differential motion cells in MT^- and additive motion cells in MT^+ (Berezovskii & Born, 2000). These cells primarily differ based on the functional

characteristics of their receptive fields. As depicted in Figure 2, differential motion cells in MT^- possess surround or side lobes of suppression (Born & Bradley, 2005; Xiao, Marcar, Raiguel, & Orban, 1997), while additive cells have no such antagonism. MT cells are selective to a specific range of speeds, stimuli sizes, and directions of motion. The antagonistic zones of differential motion cells are so named because they possess the same velocity sensitivity as the other portion of the receptive field and suppress the response of the neuron. Hence, they act like spatial differentiators in the motion domain.

Medial Superior Temporal (MST) in primates represents a functionally heterogeneous area in the extrastriate cortex that receives lateral, symmetric, and reciprocal projections directly from MT (Boussaoud, Ungerleider, & Desimone, 1990; Maunsell & van Essen, 1983). Cells in MST exhibit sensitivity to translational, spiral, rotational, expanding, and contracting motion fields, up to 100° in size (Duffy & Wurtz, 1991a, 1991b, 1995; Eifuku & Wurtz, 1998; Nelissen, Vanduffel, & Orban, 2006). Cells located in the dorsal region of MST exhibit higher sensitivity to wider motion fields, larger areal summation, and have no antagonistic surrounds compared to those located more ventrally (Nelissen et al., 2006). Differential

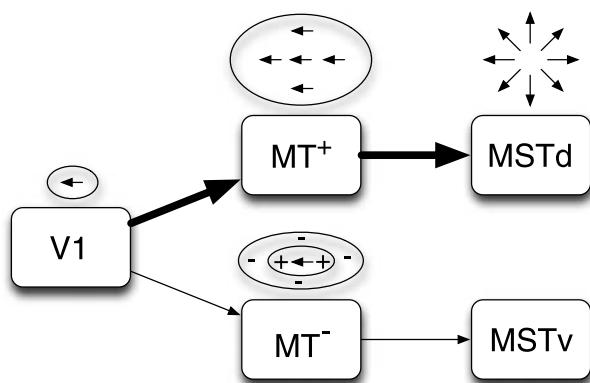


Figure 2. Proposed segregation between MT^+ /MSTd and MT^- /MSTv motion pathways based on the neurophysiological literature. The top pathway projects to MSTd where heading-sensitive cells are located via motion pooling cells in MT^+ and V1. The bottom pathway includes cells sensitive to differential motion in MT^- , with on-center/off-surround receptive field antagonism, and projects to MSTv, not MSTd. Thick arrows indicate visual areas simulated in our model to estimate heading.

motion cells from MT primarily feed ventral medial superior temporal (MSTv), while additive motion cells in MT primarily feed dorsal MST (MSTd). MSTd contains cells that are sensitive to motion patterns consistent with the estimation of translational heading whereas MSTv does not (Berezovskii & Born, 2000; Eifuku & Wurtz, 1998; Nelissen et al., 2006; Orban, 2008).

Human heading estimation models

The potential presence of both an FoE due to observer translation (defining the heading) and an FoE due to object motion (defining the object point of origin) poses challenges to both the motion pooling and differential motion theories. Royden and Hildreth (1996) and Warren and Saunders (1995) provide important data on heading perception in the presence of IMOs that tests the predictions of the two theories. Warren and Saunders found human heading bias in the direction from which independently object emanate. Royden and Hildreth found human heading bias for non-approaching IMOs in the direction of motion—opposite of that yielded in the presence of approaching IMOs. Before these data, biologically motivated models aimed to explain human heading perception in the static environment. For example, the model of Lappe and Rauschecker (1993) developed the subspace algorithm of Heeger and Jepson (1990) into a neural framework. The algorithm minimizes a residual function of five image velocity sample vectors to recover heading (Heeger & Jepson, 1990). The multi-layer neural network implementation uses the residual function form as synaptic weights between “MT” and

“MST” layers (Lappe & Rauschecker, 1993). Recent models have focused on human heading perception in the presence of IMOs. For example, a Bayesian framework that uses maximum likelihood to estimate observer translation and rotation in the presence of IMOs (Saunders & Niehorster, 2010) has been developed. The model yields results consistent with heading judgment data (Royden & Conti, 2003; Royden & Hildreth, 1996; Warren & Saunders, 1995) but is not intended to provide an explanation of how the primate brain gives rise to the heading bias.

Warren and Saunders (1995) proposed a template-matching model of human visual areas MT and MST to explain human heading bias in the presence of approaching IMOs that performed motion pooling in MT. The template-matching approach intuitively identifies how well the global pattern of optic flow experienced by the observer matches radially structured optic flow patterns or templates (Perrone & Stone, 1994). Since each template possesses a FoE from which the radial vectors emanate, and in the rigid environment without rotation, the FoE specifies the heading, the matching procedure can be used to estimate the heading. The model defines units in MST that respond to different translational optic flow fields as a function of FoE location. MST performs a Gaussian-weighted match between the velocity-sensitive MT unit activity and templates to determine the most likely focus of expansion. Since motion pooling integrates over large portions of the optic flow field, typically without regard to the presence of objects, the model predicts an averaging between the translational and object FoE that was consistent with the data (Warren & Saunders, 1995).

While the model of Warren and Saunders (1995) explains biases in the presence of approaching IMOs, Royden and Hildreth (1996) showed that it does not match the human data in the case of non-approaching IMOs (Royden, 2002; Royden & Hildreth, 1996). Royden and Hildreth instead argued that differential motion is essential to human heading perception. Rieger and Lawton (1985) developed an algorithm that used difference vectors based on the equations of Longuet-Higgins and Prazdny (1980) to compute heading in a rigid environment. Hildreth (1992) extended the difference vector approach by determining the heading of an observer in the presence of an IMO. After computing difference vectors, Hildreth defines local patches within which the center may serve as candidate FoE. By searching for evidence of vectors that lie on lines extending radially from the patch, the algorithm determines how likely the patch center is the FoE. When successful, the algorithm discounts motion due to the object, by ignoring patches of the visual field that contain inconsistent data and considering contributions in the optic flow from the translating observer. The algorithm votes across patches and determines the most likely FoE. Note that, depending on how this evidence accumulation step is performed, the algorithm can remain ambiguous whether the point is a focus of expansion or

contraction along the observer axis of translation (Royden, 1997). Royden (2002) further developed the “difference vector” model of Hildreth to include differential motion operators, which act similarly to cells found in primate visual area MT[−] with on-center/off-surround direction-of-motion antagonism. The model of Royden demonstrates human-like heading biases due to approaching and non-approaching objects, using differential motion operators inspired by cells in MT[−]. Numerous authors now consider differential motion as the best explanation for human heading perception (Duijnhouwer, Beintema, van den Berg, & Wezel, 2006; Royden, 2004; Royden & Conti, 2003; Warren, 1998).

We demonstrate that differential motion operators are not necessary to explain human heading bias data. The use of differential motion operators is difficult to reconcile with the neurophysiological data indicating that differential motion cells do not appear to project to heading-sensitive area MSTd but rather to MSTv (Berezovskii & Born, 2000; Eifuku & Wurtz, 1998; Nelissen et al., 2006; Orban, 2008). Figure 2 illustrates the pathways described in the neurophysiological literature. Mineault, Khawaja, Butts, and Pack (2012) solved an optimization problem to identify properties of expansion-selective MSTd cell subunits from MT that maximize the variance accounted for in MST cell data. Adding inhibitory surrounds to MT units does not improve the model’s ability to fit the cell data (Mineault et al., 2012), thereby supporting prior anatomical studies showing that differential motion cells do not appear to feed expansion-selective cells in MSTd. Depth discontinuities in the environment required for differential motion often do not improve heading detection thresholds (Britten, 2008; Royden & Hildreth, 1999).

Our model, based on the Visually guided Steering, Tracking, Avoidance, and Route Selection (ViSTARS) model, uses motion pooling in MT⁺ and template matching in a competitive network in MSTd to replicate the human heading biases. ViSTARS is a model of primate visual processing describing the retina–V1–MT–MST motion processing pathway. It demonstrates how MT/MST interactions can process video input for the purposes of obstacle detection, goal approach, and the estimation of heading (Browning, Grossberg, & Mingolla, 2009a, 2009b). ViSTARS is a dynamical model that explains a range of data, including the human bias demonstrated under simulating eye rotation conditions (Royden et al., 1994), and exhibits robustness to noise, but its complexity obscures the necessary conditions to explain the neural mechanisms underlying the human heading bias data (Royden & Hildreth, 1996; Warren & Kurtz, 1992; Warren & Saunders, 1995). ViSTARS unifies a number of prior models that were developed in a variety of contexts for the purposes of human navigation. For example, it integrates the FORMOTION models, which describe how V1, MT, and MST can perform motion integration and segmentation to solve the aperture problem and explain a number

of visual displays with planar motion, such as the barber pole and chopsticks illusions (Baloch & Grossberg, 1997; Grossberg, Mingolla, & Viswanathan, 2001). ViSTARS also integrates the models of Chey, Grossberg, and Mingolla (1997, 1998), which investigate how speed perception and discrimination are affected by contrast, duration, dot density, and spatial frequency (Chey et al., 1997, 1998). A related model by Pack et al. (2001) shows how areas MT⁺, MT[−], MSTv, and MSTd can interact to produce a gaze counterflow circuit to stabilize targets while performing a smooth pursuit eye movement (Pack et al., 2001). The effects that eye movements have on heading perception are also explained by the precursor to ViSTARS, the STARS model (Elder, Grossberg, & Mingolla, 2009) that uses gain fields to compensate for the effects of eye rotations.

Our present modeling work builds on ViSTARS to localize and explicate the simplest MT/MST neural circuits that explain the human heading bias data in the presence of IMOs. We do not address eye rotations in the model because eye movements did not affect the human heading bias (Royden & Hildreth, 1996; Warren & Saunders, 1995), and prior models have illustrated how larger circuits may interface with MSTd to deal with rotation (Beintema & van den Berg, 1998; Elder et al., 2009; Pack et al., 2001). Before describing our model, we first summarize the psychophysical experiments of Royden and Hildreth (1996) and Warren and Saunders (1995), then introduce the respective models that they proposed to explain these data. Our analysis shows that our model correctly replicates the direction and magnitude of heading biases at least as well as other proposed models and that the model provides a detailed and neurophysiologically consistent explanation for how the primate brain determines heading, which no other model proposed to date can do.

Heading estimation in the presence of an approaching moving object

Warren and Saunders (1995) assessed human judgments of heading in the presence of an IMO moving perpendicularly (Figure 1a, light gray dashed arrow) to the path of a translating observer (Figure 1a, black dashed arrow). Subjects viewed translational optic flow fields on a computer monitor. The optic flow stimuli were generated using randomly positioned dots on planar surfaces, where each dot moved in a manner consistent either with the background or the IMO. The object was always initially located 6° on either side of the center of the display and grew in size as the trial progressed due to the decreasing distance between the observer and the object. The object initiated movement from fixed locations and the authors altered a variable across trials called the *path angle*, reflecting the angular difference between the object and

observer FoE. Positive values indicate that the object FoE is positioned closer to the center of the screen, and the observer FoE is further to the periphery. Warren and Saunders studied path angle settings of -6° , 0° , and 6° .

Figure 3 shows sample snapshots during object motion as a function of path angle when the object begins on the right side of the display. Subjects made left-right heading judgments following each trial, relative to a “probe” location indicated by a 1° vertical line. The authors employed a two-alternative fixed choice (2AFC) experimental paradigm, and many different observer heading and object FoE cases were tested with each path angle. Sample observer and object FoE configurations are shown in **Figure 3** (Warren & Saunders, 1995). During a trial, dots appeared in their initial locations for 1 s to

communicate the beginning of the trial to the subject. Dot motion occurred for 1.5 s and dots lingered in their final positions until the subject responded (Warren & Saunders, 1995).

One of Warren and Saunders’ (1995) experimental conditions constrained the object movement to one side of the computer monitor such that it did not occlude the observer FoE on the opposing side. Under such conditions, subjects generated constant slightly positive heading biases (1.25°) toward the center of the screen for all path angles. Subjects yielded the same bias without an object present, and the authors concluded that the IMO does not impact heading judgments when it does not cross the observer FoE. This conclusion was supported by Royden and Hildreth (1996).

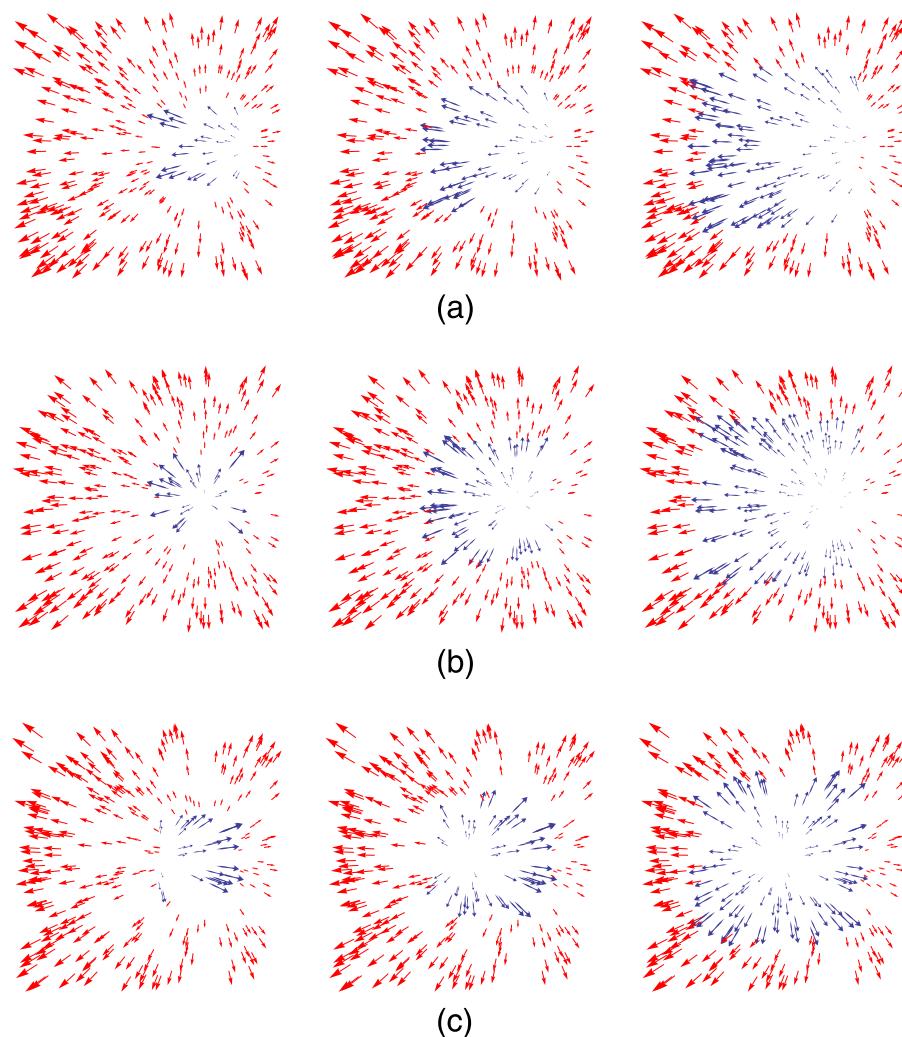


Figure 3. Illustrations of approaching IMO optic flow displays used by Warren and Saunders (1995) in the opaque object condition. Each row displays frames 1, 23, and 45 from a psychophysically presented motion sequence for a translational heading of 6.5° ; $\delta = -6^\circ$, 0° , and 6° in (a), (b), and (c), respectively, where δ denotes the *path angle* defined in the text as the difference between the observer’s and object’s foci of expansion. Background optic flow is represented in red, while the IMO is depicted by blue. In the psychophysical presentations, dots were the same color. Although the optic flow due to observer translation and object movement appear to commingle at the object boundary, the dots remained separate in the opaque object simulations.

In another experiment, the object always moved on the same side of the display as the direction of observer translation, occluding it for at least some of the trial (Warren & Saunders, 1995). Warren and Saunders postulated that human heading judgments could be impacted either by the object obscuring the observer's FoE with inconsistent motion or due to the fact that the observer FoE is not visible for some portion of the trial. To disambiguate these possibilities, Warren and Saunders employed three object types that varied in their opacity. In the opaque object case depicted in Figure 3, the background dots are suppressed in areas that the object obscures (i.e., the object occludes the background). In the transparent object case, the dots belonging to the object and background field coexist intermingled (i.e., the object does not occlude the background). Finally, the "black" object case features no dots where the object exists. Warren and Saunders (1995) reported strong positive biases under the opaque (6°) and transparent (4°) object conditions when the path angle was set to 6° . In other words, when the object approaches the observer from closer to the center of the screen than the observer FoE and the object occludes the translational FoE, subjects experience strong biases toward the center of the screen. When the path angle was set to -6° , the authors found a stronger negative bias for the opaque object (-2°) compared to the transparent object (-0.5°). In other words, when the object approaches the observer from the edge of the screen and occludes the observer FoE, subjects experience heading biases in the direction of the edge, albeit weaker than when it approaches from the center of the screen. When the path angle was set to 0° , both opaque and transparent object conditions produced approximately equivalent positive heading biases of 2° , similar to those generated when the object did not cross the observer FoE. The black object conditions yielded a small positive bias under all path angle conditions approximately equivalent to the bias yielded in the absence of an IMO ($<2^\circ$; Warren & Saunders, 1995). The black object results suggest that an inability to see the observer FoE alone does not induce a heading bias, but

when combined with dot motion from the object, error is introduced into human heading judgments.

Heading estimation in the presence of a moving object maintaining a fixed distance

Royden and Hildreth (1996) studied human heading accuracy for objects that maintain a fixed distance from the observer (Figure 1a, dashed line). Subjects viewed translational optic flow fields on a computer display represented by dots moving on two fronto-parallel depth planes. One experiment assessed heading judgments to horizontal and vertical object motion. Four horizontal headings of 4° , 5° , 6° , and 7° were simulated on the right side of the display, and vertical headings of 0° and 2° above and below the horizontal midline of the display were tested (Royden & Hildreth, 1996). The object was opaque, possessed denser dot motion than that of the surrounding translational field, and moved with a constant speed either right, left, up, or down. For the horizontal movement conditions, the object started at one of six different starting locations. Because the object maintained a fixed distance with respect to the observer, it only appeared to move horizontally or vertically during trials. Figure 4 depicts snapshots during different illustrative motion sequences. Subjects viewed the initial frame of the sequence, then initiated each trial via a button press. Dots remained in their final locations until the subject placed the mouse cursor in the perceived direction of motion and clicked to conclude the trial (Royden & Hildreth, 1996).

When the object moved vertically, Royden and Hildreth reported average horizontal heading biases as a function of horizontal starting position, averaged across all subjects. Positive biases and starting positions correspond to the right side of the screen, whereas negative values correspond to the left side of the screen. This definition of bias differs from that of Warren and Saunders (1995), who defined subject heading bias relative to the object FoE. In both upward and downward moving cases where object

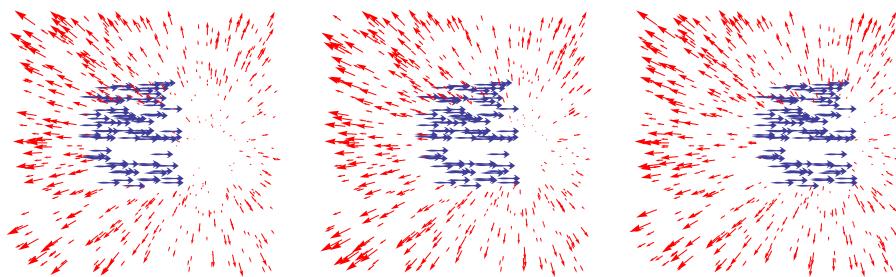


Figure 4. Sample fixed-distance IMO optic flow display used by Royden and Hildreth (1996). Frames 1, 10, and 20 from the rightward object sequence R3 are displayed (see [The model](#) section for details). At the trial outset (left), the object (blue) does not occlude the observer FoE (7°) in the background optic flow field (red). By the final frame, the IMO completely occludes the observer FoE.

motion occluded the observer FoE for less than 50% of the trial, subjects produced a bias of approximately zero, similar to results garnered by Warren and Saunders (1995) under analogous conditions. However, when the object occluded the observer FoE for at least 50% of the trial, a small negative, leftward bias of approximately -0.5° occurred irrespective of the vertical direction of object motion.

When the object moved horizontally and occluded the observer's FoE for more than 50% of the trial, Royden and Hildreth (1996) found different directions of bias depending on the direction of object motion. Conditions in which the IMO did not obscure the FoE resulted in virtually no heading bias. Subjects reported negative average biases (approximate maximum magnitude of -1°) to leftward moving objects that occluded the observer FoE for part of the trial (Royden & Hildreth, 1996). In the rightward moving object condition, there were positive rightward biases (approximate maximum

magnitude of 0.5°) when the object crossed the observer FoE for part of the trial. Therefore, when laterally moving objects obscure the observer FoE for at least some portion of the trial, human heading judgments become biased in the direction of the object motion. The direction of heading errors found by Royden and Hildreth represents the opposite of that found by Warren and Saunders (1995). Objects that maintain a fixed distance from the observer affect heading estimation differently from those that approach.

In this article, we present a neural model of primate visual areas MT⁺ and MSTd. We use this model to unify the psychophysical data on approaching and non-approaching IMOs. As shown in Figure 5, our model pools motion over V1 motion representations in model MT⁺ and performs template matching in a competitive network in model MSTd, maintaining consistency with the primate neurophysiological data and demonstrates human-like heading bias. Our model explains the psychophysical

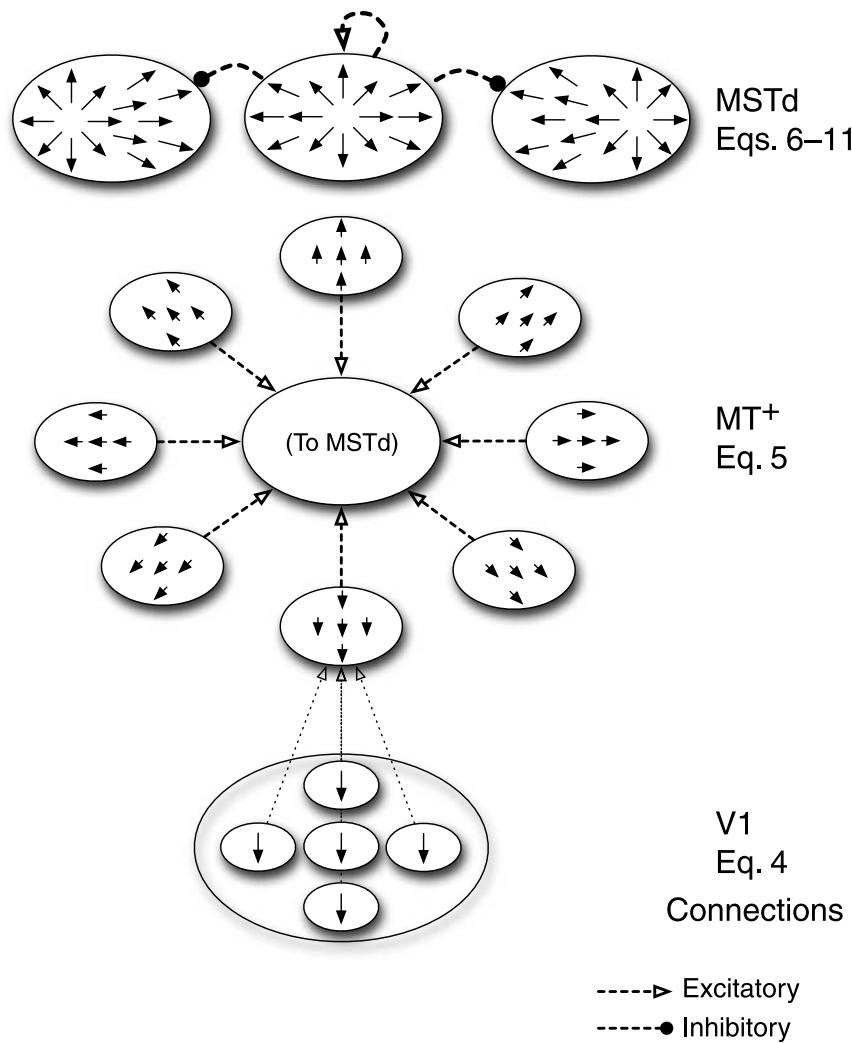


Figure 5. Model diagram. Analytic representations of the input sequence are computed in model area V1. Model area MT⁺ pools over V1 cell responses, which feed template matching in model area MSTd. Model MSTd cell responses are smoothed in a heading matching layer and MST units compete over time. The maximally active unit represents the best match and is taken as the heading.

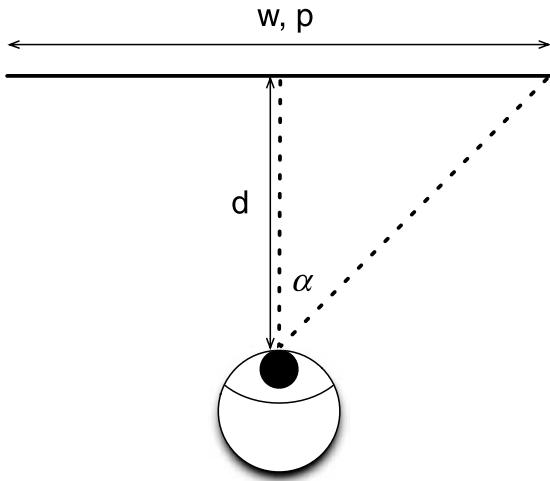


Figure 6. Experimental setup. A model observer viewing α° of optic flow fields from Royden and Hildreth (1996) and Warren and Saunders (1995) at a distance d cm from the monitor p pixels and w cm wide (top-down view).

data through an emergent *peak shift* in model MSTd (Figure 6).

The model

We replicate the displays shown to human subjects participating in the psychophysics studies of Royden and Hildreth (1996) and Warren and Saunders (1995). The subjects viewed optic flow displays on a monitor p pixels and w cm wide at a distance d cm. The studies report experimental conditions and results with respect to degrees of visual angle. We convert α degrees to P pixels and vice versa using

$$P = \frac{pd}{w} \tan\left(\frac{\alpha\pi}{180}\right), \quad (2)$$

$$\alpha = \frac{180}{\pi} \arctan\left(\frac{wP}{pd}\right). \quad (3)$$

The study of Royden and Hildreth (1996) employed an Apple 21" monitor paired with an Apple Quadra 950 workstation. We assume that the study used a 21" Macintosh Color Display (19" viewable area) with 1152 × 870 pixel resolution with an approximate physical aspect ratio of 1.06,¹ as was standard with this type of Apple computer at that time. The viewing distance of the subjects is 30 cm. Using Equation 3, we find that the 30° × 30° viewing window and 10° × 10° IMO are 529 × 423 pixels and 173 × 137 pixels, respectively.² Royden and Hildreth generated optic flow stimuli using random

dots refreshed on the monitor at 25 frames/s; each stimulus had a duration of 0.8 s, for a total of 20 frames per trial. Royden and Hildreth used dot densities of 0.56 dot/deg² and 0.8 dot/deg² for the background and object, so we generate our backgrounds and objects with 500 dots and 80 dots, respectively. Adopting the convention used in the study, the center of the viewing window represents the origin of the image plane. As such, negative and positive positions reflect those to the left and right of the center, respectively. By Equation 3, the simulated horizontal observer headings of 4°, 5°, 6°, and 7° become 69 px, 86 px, 104 px, and 121 px, respectively. In the leftward motion conditions, the IMO moved from -1.4°, 0.6°, 4.7°, 8.7°, 10.7°, and 12.7° to -7.88°, -5.88°, -1.78°, 2.22°, 4.22°, and 6.22°, respectively. In the rightward motion conditions, the IMO moved from -9.9°, -5.9°, -1.9°, 0.2°, 2.2°, and 6.3° to -3.42°, 0.58°, 4.58°, 6.68°, 8.68°, and 12.78°, respectively. We designate conditions in which the object moves left and right with an "L" and "R," respectively, and append ascending numbers to reflect the relative starting position of the object. For example, in condition L1, the object began further to the left than in L3. The initial and final positions of the IMO replicate those used in Royden and Hildreth. While in motion, the object moves with a constant velocity to the end point, as described in Royden and Hildreth.

Warren and Saunders (1995) used a 1280 × 1024 px monitor with a 60-Hz refresh rate. Since the viewing window for the visual displays subtended 40° × 32° and were viewed from a distance of 43 cm, we find that, according to Equation 3, this is equal to 961 × 757 pixels. Similar to Royden and Hildreth (1996), Warren and Saunders used 10° × 10° objects that are equivalent to 231 × 131 pixels on the display. Each stimulus was presented for 1.5 s and was rendered at 30 frames/s, resulting in a total of 45 frames per trial. We simulate the opaque, transparent, and black approaching IMO cases tested by Warren and Saunders. The background consisted of 300 dots, while the object had 25 dots initially. Because both the observer and object translate at a constant speed, the approaching object grows linearly in time across frames. The object initially subtended 10° × 10° and grew to about 20° × 20° by the end of the trial. As described by Warren and Saunders, we fix the initial position of the object to ±6° (±137 px) from the center and constrained the object movement such that it begins on the same side as the simulated heading. The headings were 0°, ±2°, ±3–11° in 0.5° increments, ±12°, and ±14°. The object motion remained fixed, and the path angle δ varied between -6°, 0°, and 6° (i.e., the angle between the object FoE and that of the observer). If objects in either set of psychophysics experimental conditions grew or moved beyond the viewing window, we clipped the object at the viewing boundaries.

Our simulations were performed on a 2.66-GHz 8-core Apple Mac Pro with 16-GB RAM in Wolfram

Mathematica 7. We implemented a simplified version of the ViSTARS heading model (Browning et al., 2009b), focused on the core computations involved in heading estimation.

Model V1

In Model V1, we analytically compute first-order representations of the optic flow field in the non-approaching and approaching IMO conditions. Using Equation 1, we set $\vec{r} = 0$ and t_y to zero, since the observer only translates in depth and there is no rotational optic flow in the displays reported in the psychophysical studies. Hence, Equation 1 reduces to

$$I_l(x, y) := \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} x t_z - t_x \\ y t_z \end{pmatrix}. \quad (4)$$

In Equation 4, (\dot{x}, \dot{y}) represent the horizontal and vertical flow components at the position (x, y) , t_z signifies the depth component of the translational velocity of the observer, t_x indicates the horizontal component of the observer translation, Z is the distance from the observer to the point in space represented by the dot, and l specifies the l th frame in the motion sequence. Although Equation 4 only considers first-order information (i.e., the velocity vector field over time), the velocity field representation of optic flow can yield the same heading estimates in humans as fields containing higher order information for dot displays (Warren et al., 1991). While we and many models assume velocity field representations of optic flow, higher order flow may afford amenable information for navigation to the observer. We use the notation $I_l(x, y)$ to represent the vector-valued optic flow field (\dot{x}, \dot{y}) at frame l with spatial location (x, y) . In the non-approaching condition, we generate optic flow fields using Equation 4 with $t_z = 200$ cm/s, the translational speed of the observer toward the background dot planes located at 400 cm and 1000 cm (Royden & Hildreth, 1996). Each object point moved at a constant speed of 8.1 deg/s, as described by

Royden and Hildreth (1996). In the approaching IMO case, we reproduce the 5-s time to contact between the observer and fronto-parallel dot plane by setting $Z = 1000$ cm and $t_z = 200$ cm/s. Where the object exists on the display, we set $t_z = 300$ cm/s to recreate the 3.33-s time to contact between the observer and the object. We consider the opaque object experimental condition of Warren and Saunders (1995), as it yielded the most pronounced heading biases. In order to generate the opaque object, we replace background points in the object region with those corresponding to the object. Figures 3 and 4 exhibit sample V1 representations used in the simulations.

After converting the degrees of visual angle subtended by the optic flow displays in each respective study into pixels (Equation 3), we generate uniformly sampled heading templates $T_i(x, y)$ by substituting pixel locations (x, y) into Equation 4 and normalizing each vector to unit length, with FoE at each horizontal position $i = t_x$. Figure 7 shows sample templates used in the simulations.

Model MT⁺

Model V1 projects directly to Model MT⁺, where cells have receptive fields that integrate over short-range motion of particular velocities. We define the pooled MT motion $D_l(x, y)$ according to

$$D_l(x, y) := (I_l * G_{MT})(x, y; \mu_{MT}, \sum_{MT}, r_{MT}), \quad (5)$$

where $*$ denotes the 2D convolution operator, G_{MT} is a 2D discrete multivariate Gaussian kernel with mean μ_{MT} and covariance matrix \sum_{MT} normalized such that all points in the kernel's support sum to unity, and r_{MT} defines the kernel radius. We model MT⁺ cells with circular receptive fields; hence, we set $\mu = \bar{0}$ and \sum_{MT} such that $\sigma_x = \sigma_y = \sigma_{MT}$ and the covariance ρ between x and y is zero. We employ a single parameter set to conservatively simulate MT⁺ cell receptive field properties found in neurophysiological studies (Born & Bradley, 2005; Churchland et al., 2005; Nelissen et al., 2006)

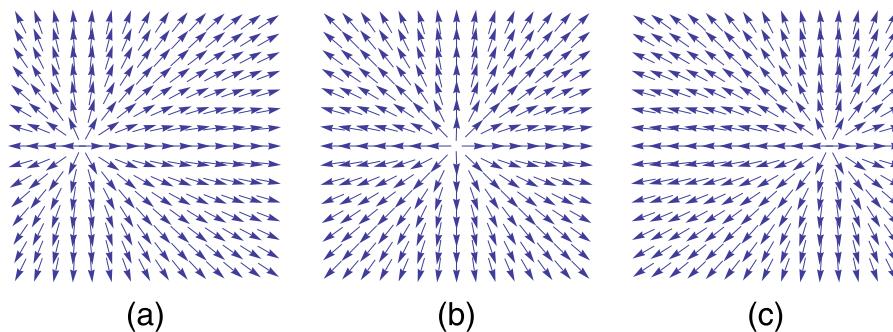


Figure 7. Sample heading templates. From left to right, we display -8° , 0° , and 8° normalized templates used in the non-approaching object simulations ($30^\circ \times 30^\circ$ display). Note that 0° refers to the center of the display and positive and negative angles correspond to the FoE position on the right and left sides of the display, respectively.

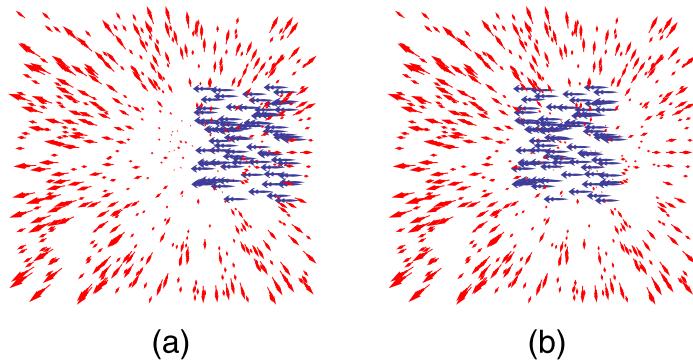


Figure 8. Sample MT representations after motion pooling. Panels (a) and (b) show frames 1 and 20 of the non-approaching IMO condition L4, respectively.

found strong fMRI responses to kinetic gratings compared to baseline conditions in area MT for spatial periods of 0.125 deg/cycle and to random textured patterns 3–28° in diameter, with responses increasing with size. We used $\sigma_{MT} = 0.05^\circ$ and $r_{MT} = 3^\circ$ to fit these findings. Figure 8 shows sample frames from the motion sequences after pooling. All convolutions in MT^+ are zero-padded and performed component-wise.

Model MSTd

In Model MSTd, we perform a template match between optic flow frames $D_l(x, y)$ and all templates $T_i(x, y)$. That is, for a given $D_l(x, y)$, we match against $T_i(x, y)$ for all horizontal headings i . We obtain a scalar value p_i^l for each frame l at the horizontal heading i , representing the cosine similarity (i.e., inner product) between distance-weighted vectors in the motion frame and those in the template defined by

$$p_i^l = \lambda \sum_{\{x,y\}} W_i(x, y) \left(\sum_{\{x,y\}} \frac{T_i(x, y) D_l(x, y)}{\|D_l(x, y)\|_2} \right). \quad (6)$$

In Equation 6, $W_i(x, y)$ represents a distance-dependent weighting from horizontal heading i . We use inverse 2D Euclidean distance, scaled by a parameter λ to adjust the spatial extent of the templates. We selected $\lambda = 300$ for a broad spatial tuning. The inner summation performs component-wise multiplication between vectors in template $T_i(x, y)$ and MT^+ output $D_l(x, y)$ for every spatial location and frame. The resulting vector is normalized by the L^2 (Euclidean) norm (denoted $\|D_l(x, y)\|_2$) and then the vector components $\{\dot{x}, \dot{y}\}$ are summed.

We subsequently smoothed the 1D pattern match distribution in MSTd according to

$$P_i^l := (p_i^l * G_{MST})(i; \mu_{MST}, \sigma_{MST}, r_{MST}), \quad (7)$$

where $*$ is 1D cyclic convolution and G_{MST} is a normalized 1D Gaussian kernel. We set the radius r_{MST} to 12°, $\sigma_{MST} = 2^\circ$, and mean of the MST kernel $G_{MST} \mu_{MST} = \bar{0}$. These parameters conservatively mimic neurophysiological studies reporting greater areal summation in primate area MSTd compared to MT (Duffy & Wurtz, 1995; Nelissen et al., 2006).

Finally, we introduce a dynamical competitive network in MSTd. Grossberg (1973) analyzes the following network equation, termed a *recurrent competitive field*:

$$\dot{x}_i = -Ax_i + (B - x_i)(f(x_i) + I_i) - x_i \sum_{k \neq i} f(x_k). \quad (8)$$

In Equation 8, A specifies the passive decay rate, B defines the saturation upper bound, $f(w)$ describes the signal function, and I_i defines the external input to unit x_i . The *signal function* $f(w)$ dynamically specifies the nature of the feedback a cell receives relative to its current activity. We solve Equation 8 at equilibrium (i.e., $x_i = 0$) and use a faster-than-linear signal function $f(w) = w^2$ to form a choice, or winner-take-all, network (Grossberg, 1973). We obtain recurrent MSTd units, M_i^l , after substituting in the smoothed pattern match distribution, P_i^l , and setting $A = 1$, $B = 1$:

$$M_i^l = \frac{g(P_i^l)^2}{1 + \sum_k g(P_i^k)^2}. \quad (9)$$

Because the minus sign before the summation term $-x_i \sum_{k \neq i} f(x_k)$ in Equation 8 occurs before the factor x_i , the inhibitory effect cell k has on cell i is, at equilibrium, divisive rather than subtractive, as shown in Equation 9. The function $g(x_i)$ is defined as a linear accumulation of network activity between stimulus frames:

$$g(x_i) = c g(x_{i-1}) + (1 - c)x_i. \quad (10)$$

Hence, with Equations 9 and 10, we accumulate the smoothed pattern match over time. We choose $c = 0.3$ to temporally weight network activity due to new visual information higher than that of recent history. Following MSTd competition, we determine the judged heading direction by selecting the template that has the most activation in the final frame in the motion sequence. The judged heading i^* is found by the following equation at the last frame n :

$$i^* = \operatorname{argmax}_i M_i^n. \quad (11)$$

Finally, we determine the heading bias of the model by subtracting the judged heading i^* from that generated by the network in the absence of an IMO. Because Royden

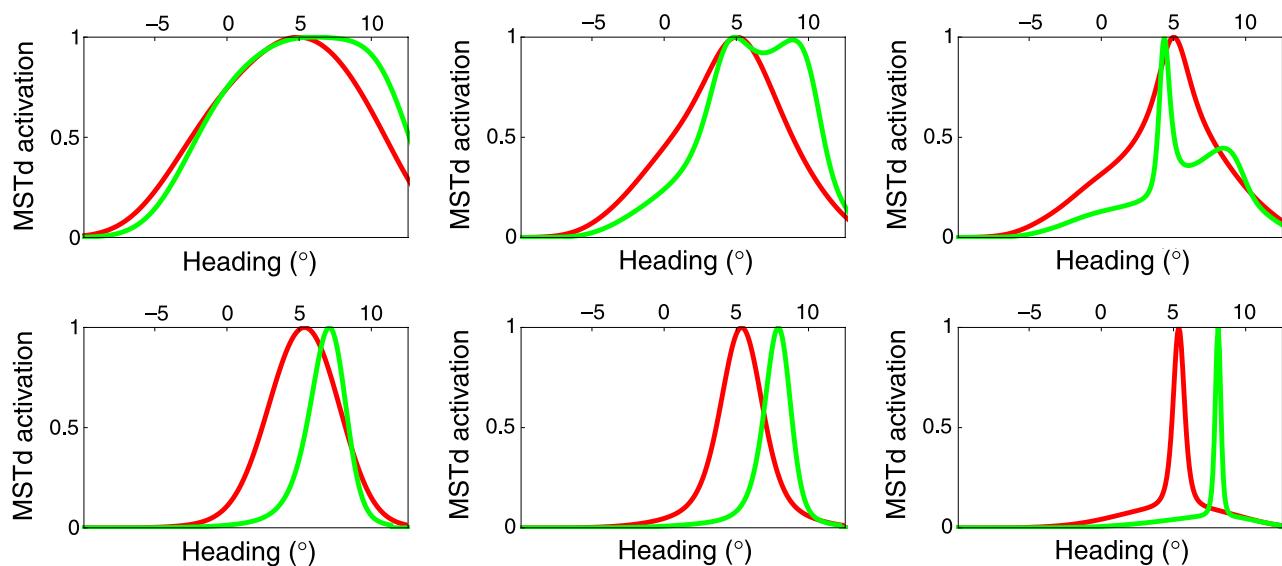


Figure 9. MSTd responses at different times during the presentation of random dot display sequences. Red depicts activity in the absence of an IMO, whereas green curves show the time-averaged responses with the IMO present. The top row shows the 5th, 10th, and 20th frame from left to right of the non-approaching object sequence simulating the Royden L4 condition. The observer heading is 5° to the right and a bias of -0.75° is generated in this trial. The second row shows sample MSTd responses in the approaching IMO condition, showing frames 10, 20, and 45. The observer heading is 5.5°, while the object FoE is 11.5°. The bias is 2.7° to the right, away from the center of the screen.

and Hildreth (1996) define the sign of the bias differently than that of Warren and Saunders (1995), we employ the conventions used in each respective study when reporting the model results. Hence, in the non-approaching object condition, we define a positive bias as a heading estimation too far to the right of the screen. In the approaching object condition, we define a positive bias as a heading estimation too far toward the center of the screen. Figure 9 displays some sample MSTd responses along with their biases. Due to the random selection of dots in each experimental condition, we ran each configuration 10 times and averaged to obtain the reported headings.

Results

Figures 10a and 10c depict the average model performance when viewing the non-approaching IMO compared against the data from Royden and Hildreth (1996), adapted from Royden (2002). Figures 10b and 10d are adopted from Figure 6 of Royden, showing the model performance of Royden compared against the same human data in Figures 10a and 10c. Royden and Hildreth averaged across subjects and observer headings, while we average over observer headings and perform 10 trials per heading. Although our model returns deterministic results given repeated presentations of the same dot motion sequence, the displays consisted of random dot patterns, which introduced intertrial variation. Our model yields similar bias curves to those found by Royden and

Hildreth. Across leftward and rightward IMO conditions, we consistently obtained biases in the direction of the object motion: left for a leftward IMO and right for a rightward IMO. We obtained $r = 0.94$ and $r = 0.71$ using Pearson's correlation when comparing our performance against the human data for the leftward and rightward IMO conditions, respectively. Similarly to the psychophysics results, the model produces the largest error when the IMO occludes the observer heading. When the IMO does not, or hardly, covers the observer heading during the trial, the bias is small.

Figure 11 displays the model performance in the approaching IMO condition averaged across observer and object headings for a given path angle δ . Biases averaged across subjects and observer and object FoE reported by Warren and Saunders (1995) are also drawn; these values are approximated based on Figure 4 from Warren and Saunders. The model fits Warren and Saunders' data well, with $r = 0.99$, $r = 0.98$, and $r = 0.86$ in the opaque, transparent, and black object conditions, respectively. These results demonstrate that approaching IMOs result in biases in the direction of the object FoE (Warren & Saunders, 1995).

Discussion

We have presented a motion pooling model of MT⁺ and MSTd that explains the heading biases produced in the psychophysical studies of Royden and Hildreth (1996) and

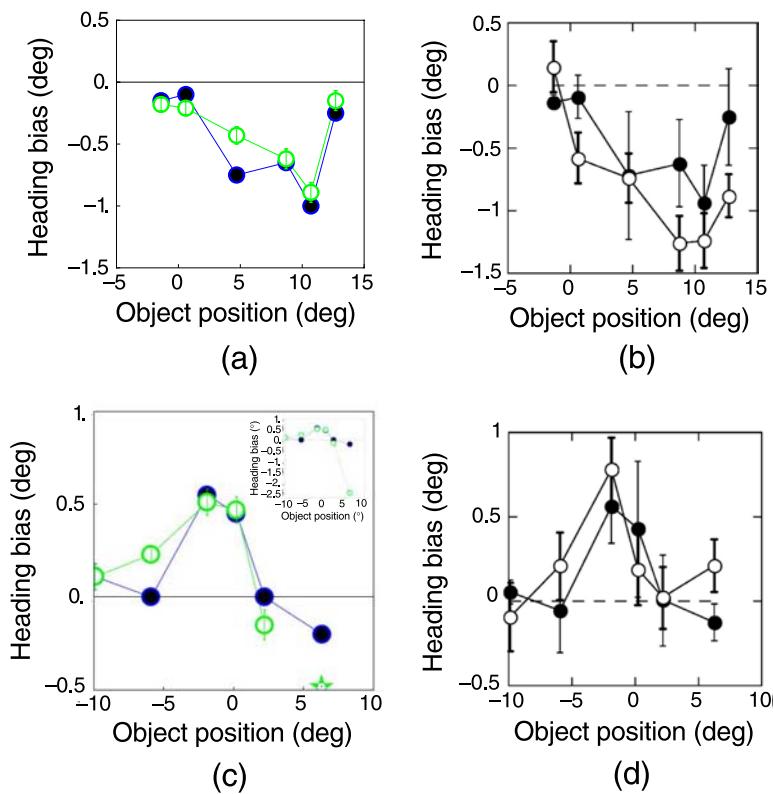


Figure 10. Simulated biases averaged across different heading conditions for the non-approaching IMO. Panels (a) and (c) show the simulation results (green) and human psychophysical data (blue) for leftward and rightward object motion, respectively. Model results for condition R6 ($\bar{x} = -2.46^\circ$) is shown in the inset of (c). Using Pearson's correlation, we found $r = 0.94$ and $r = 0.71$ for the left and right conditions, respectively. Human subject biases were approximated from Figure 4 of Royden and Hildreth (1996). Error bars indicate 1 SEM. Panels (b) and (d) show the performance of the model of Royden (2002; white) using asymmetric differential operators compared against that yielded by humans (black).

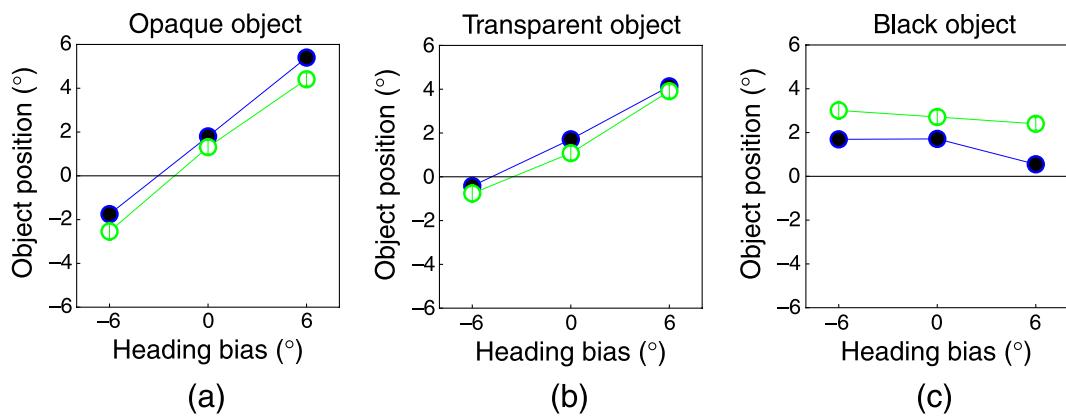


Figure 11. A comparison between average subjects psychophysical data in Warren and Saunders (1995; blue) and the average present model results (green) as a function of path angle (denoted δ) for the approaching IMO condition. Subject biases were approximated from Figure 4 of Warren and Saunders. Error bars indicate 1 SEM. (a) Model findings closely reflect those reported by Warren and Saunders, indicating that in the presence of an IMO occluding the observer's FoE, heading biases occur in the direction of the object FoE. Using Pearson's correlation, we find $r = 0.99$. (b) Model results also closely match the human data when the approaching moving object is transparent ($r = 0.98$). (c) Simulation results compared to the human heading data in the black object condition ($r = 0.86$).

Warren and Saunders (1995). As found in humans by Royden and Hildreth, when non-approaching objects cover the FoE of a translating observer, the model produces biases in the direction of object motion. By contrast, when the object approaches the translating observer and covers the observer FoE, the model generates biases in the direction of the object FoE. Our model unifies the results of both studies while remaining consistent with known neurophysiology. The model also indicates that the primate visual system can determine the direction of heading by pooling motion in MT⁺ and competition in MSTd, without needing units sensitive to differential motion. As depicted in Figure 9, our model explains the two sets of data using a *peak shift* in the MSTd unit distributions. In the non-approaching case, the peak shift occurs because when the IMO occludes the observer's FoE, the motion around the FoE is inconsistent with a heading in that direction. The MSTd population distribution thus has a trough around the position of the IMO, which in turn causes the peak to split into a bimodal distribution with a maximum peak on one side of the IMO. In the approaching case, the peak shift occurs due to the MSTd distributions corresponding to the observer's and object's FoE being close enough that they merge and produce a peak in between the FoE positions. Royden (2002) notes that her model based on MT⁻ differential motion cells replicates the human heading results without actively removing the IMO as would other models, such as Hildreth (1992). Our model also replicates the human data without removing the IMO to compute heading, and because it does not require differential operators, we claim that it is more consistent with neurophysiological data.

Approaching IMO

When the approaching IMO occluded the observer's FoE, the optic flow pattern attributed to the IMO progressively became more influential during the trial due to the increase in size of the IMO's representation in model area V1. Peak activities in MSTd units reflected this trend by beginning closer to the observer's FoE and over time shifting toward the IMO's FoE. The fit we obtained in the approaching IMO simulations ($r = 0.99$ opaque IMO, $r = 0.98$ transparent IMO, $r = 0.86$ black IMO) was not surprising, since the model of Warren and Saunders (1995) pools motion and also explains these data. Since the black object had no dot motion defined within its boundary, we obtain a relatively flat heading bias curve as a function of path angle, also mimicking the decrease in bias as path angle increased, seen in the human data. The bias curve in the black object condition is not flat because of intertrial variation and positional effects attributed to the object always beginning $\pm 6^\circ$ from the center of the display. We discovered relative speed of the object compared to that of observer translation and the amount of motion pooling in MT⁺ altered the model

performance. For a given path angle, increasing the speed of the approaching object tended to globally shift the biases produced for all tested observer and object FoE pairs. This is because pooling locally disperses motion direction contributions, consequently increasing neighboring template-match scores. Adjusting the amount of MT⁺ pooling (r_{MT} , Σ_{MT} ; Equation 5) had large effects and influenced each observer and object FoE pair based on the path angle and context. Similarly, as Σ_{MT} adjusts the model MT⁺ cell spatial integration extent, this parameter may largely shift the MSTd cell template-match scores. By virtue of the constraint on the visual displays that the observer heading be on the same side as the approaching object, the motion pooling and template-match distributions in MSTd were usually unimodal due to smoothing that merges proximal match activity. Network accumulation in MSTd (c ; Equation 10) also impacted heading biases by influencing the temporal sensitivity of match scores. Our selection of c allowed the network to integrate information over time but not disregard the recent past. Less smoothing in MSTd (σ_{MST}) also increased the network sensitivity to peak shifts and to other changes in the match scores between frames. Furthermore, we observed an expected symmetry of trials conducted on the right and left sides of the screen. That is, if we reflected each frame of an approaching moving object sequence about the center of the screen, we obtained the same biases. This is not true of the non-approaching object due to the lack of positional symmetry in the design of the study (Royden & Hildreth, 1996). Interestingly, biases remained insensitive to a variety of dot densities, echoing the findings of Warren and Hannon (1990) that dot density did not impact percent correct performance in their 2AFC paradigm.

Although approaching IMOs have been studied in the literature, we do not know of any thorough investigations of receding IMOs. Figures 12a and 12b show the first and last frames of a receding IMO sequence. Our model predicts heading bias in the direction opposite of the FoC relative to the observer FoE position. Our analysis indicates that differential motion models, such as that of Royden (2002), may not make the same prediction because the direction of bias is dependent on the relative speed between the IMO and background. For example, if the receding object speed is much greater than that of the background (perhaps similar to the right side of the IMO in Figure 12a), the object vectors dominate in the vector subtraction and the heading estimate will be biased toward the object.

Non-approaching IMO

MSTd template-match distributions in the non-approaching condition were often bimodal, due to “good” matches immediately around the discontinuities between the object and the background (i.e., motion boundaries). The activation within the object boundaries was reduced

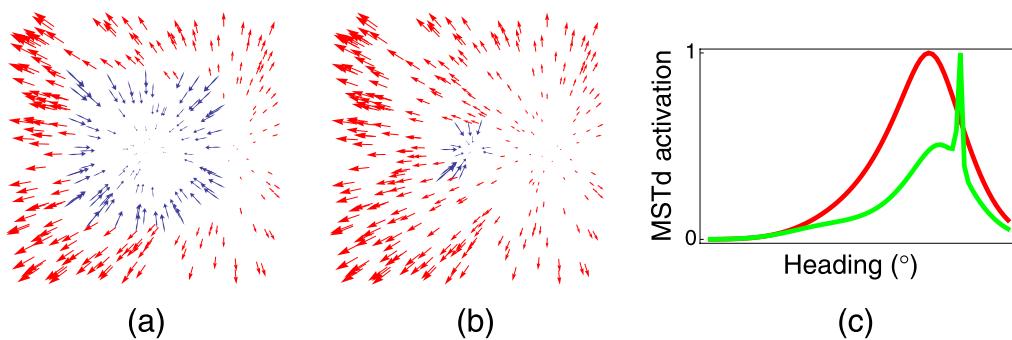


Figure 12. Model response to a receding IMO. (a) The first frame in the sequence. (b) The last frame in the sequence. (c) The response to the receding IMO in Model MSTd. Our model predicts a heading bias in the direction opposite of the FoC relative to the observer FoE position (green) compared to when no object is present (red).

because MSTd units obtain suboptimal pattern matches when sampling within the IMO's extent and become suppressed in the competition. Depending on the amount of pooling, the proportion of the trial that the object occludes the observer FoE, and the amount of competition in MSTd, the heading that gives rise to the “surviving” peak in MSTd may change. For example, strong competition magnifies small differences between both peaks because the recurrent competition field with a faster-than-linear signal function must make a choice (Grossberg, 1973). Additionally, before or after the IMO passes over the observer FoE, weaker competition can either expedite the dominance of an emerging peak or stronger competition can prolong the dominance of an existing peak. We use distance-dependent weighting to help fit the human data. This is unlike the model of Royden (2002) that relies on distance-dependent weighting to prevent the network from producing biases when the IMO is positioned far away from the translational FoE. The recurrent competitive field and motion accumulation after the MSTd pattern match preclude such biases in our model because the object would produce relatively low match scores compared to the visible translational FoE and hence lose the competition over time (Grossberg, 1973; Royden, 2002). Our model response to the R6 condition did not fit the magnitude of heading bias reported experimentally, although our model, unlike the model of Royden, matched the direction of bias. We note that the model of Royden also deviated on the R6 condition, producing a bias of the opposite sign of that reported by Royden and Hildreth (1996). During our analysis, we were able to obtain a better fit to Royden and Hildreth's data using a different set of parameters ($r = 0.86$); however, with these parameters, the fit to Warren and Saunders' (1995) data was reduced. Parameters in the present study were chosen to match the psychophysical and neurophysiological data with the minimal number of parameters and to utilize a single set of values across all our simulations. This reduces the model complexity and allows for greater understanding of the computations taking place. Ongoing research is investigating how multiple sets of receptive fields may interact within MT and MST and how best to

parameterize them within the model. Although the results of Royden and Hildreth indicate human heading bias in the direction of object motion when the observer's FoE is occluded, the population vector of monkey MSTd cell responses may only reflect heading error if the object motion greatly deviates from that of the surrounding optic flow produced by the observer translation (Georgopoulos, Schwartz, & Kettner, 1986; Logan & Duffy, 2006). In other words, MSTd cells in monkey may only yield a biased representation in a subset of the non-approaching IMO cases tested in this article. Our model can account for these differences with a change in parameters.

Optic flow illusion

Superimposing fields of radially expanding and laterally moving dots, often on different depth planes, forms an effect known in the literature as the optic flow illusion (OFI; Duffy & Wurtz, 1993). The perceived FoE shifts in the direction of the lateral dot motion proportional to its speed (Pack & Mingolla, 1998; Royden & Conti, 2003). Royden and Conti (2003) claim that this supports the hypothesis that the visual system performs local differential motion. Numerous manipulations, such as superimposing two radial fields, separating the fields by a gap, and changing the lateral field to a rotating field, have also been investigated (Duijnhouwer et al., 2006; Duijnhouwer, Van Wezel, & van den Berg, 2008; Royden & Conti, 2003). The superposition of two radial fields may be interpreted as laterally sliding the closer plane uniformly as the observer translates straight ahead, which shifts the perceived FoE in the direction opposite of the planar movement. Our model shows the same direction of bias as the psychophysics data on the two superimposed radial field case (Royden & Conti, 2003). In the simplified form described here, our model cannot account for the original OFI because full field lateral dot movement induces zero bias for slow lateral dot speeds and a slight bias in the direction opposite of the dot motion direction for faster speeds. However, we argue that the OFI may arise due to global motion integration, since the context may reflect

that of visual stability during a smooth pursuit eye movement (Duffy & Wurtz, 1993; Pack & Mingolla, 1998). This context significantly differs from that of the IMOs analyzed in the present article, wherein optic flow due to object translation surrounds that of the IMO, and the visual system may, therefore, use different visual circuits. We believe that adding a smooth pursuit counterflow stage, such as that proposed by Pack et al. (2001), multiplicatively combining retinal and extraretinal signals (Beintema & van den Berg, 1998), or using gain fields (Elder et al., 2009) in the model would allow the model to account for the OFI human data (Royden & Conti, 2003). Whether counterflow alone is sufficient to explain the human data on the original OFI, or its modifications, will be dependent on the parameterization of the modified model.

Timing

We assume that heading judgments in the model are made after the final frame is presented. Analogously, this assumes that subjects exclusively decide on their direction of heading after viewing the information present in the final frame of motion, but humans may decide on a heading at any point during the trial and may ignore some of the available information. In fact, Royden (2002) observes that the psychophysical data reported here fit her model better earlier in a trial. In the approaching moving object condition, it would seem plausible if humans valued early information due to the object expansion that progressively obscures the optic flow field (Royden, 2002). When the object does not approach the observer, one can imagine later frames of the display providing more reliable information about the translational FoE if the object initially obscured but later moved away to reveal it (Lappe & Krekelberg, 1998). Royden and Hildreth (1996) and Warren and Saunders (1995) employed similar presentation protocols as summarized in the **Introduction** section, with the exception of the differing means of response (cursor clicks compared to left-right judgments) and the fact that subjects in Royden and Hildreth clicked a button to initiate the trial whereas in Warren and Saunders the trial began automatically. Our model currently samples the visual field uniformly when performing template matches; however, neurophysiological evidence that more cells in MST may have a more peripheral preferred azimuth of FoE but exhibit a greater sensitivity foveally (Gu, Fetsch, Adeyemo, DeAngelis, & Angelaki, 2010) exists. In a study requiring eye fixation, such as Royden and Hildreth, template sampling and weighting differences may change the pattern of results.

MSTd cell types

At present, our model does not include the full variety, or complexity, of cells found in primate MSTd. In order to

keep the model simple and directed toward the assessment of self-motion, we have focused on radially expansive cells with receptive fields that cover most, if not all, of the visual field. No additional properties of MSTd cells were required to explain the human data discussed herein. Evidence suggests that a number of MSTd cells that exhibit sensitivity to radial expansion not only respond in the context of self-motion but also to aspects of object motion in the scene. The response of individual MSTd neurons may be a complex combination of local object and global motion (Sato, Kishore, Page, & Duffy, 2010). While the response to object motion in MSTd is related to the work we present here, it does not seem necessary to explain human heading biases in the presence of IMOs. Although MST cells of differing heading direction preferences in our model inhibit each other via the term $-x_i \sum_{k \neq i} f(x_k)$ of **Equation 8**, which has a minus sign, the effect is neither global nor local subtraction (Royden, 2004). Our model uses divisive rather than subtractive normalization, as can most readily be seen in **Equation 9** due to the shunting inhibition by x_i in **Equation 8**, which has different effects than subtraction (Grossberg, 1973; Heeger, 1992; Levine & Grossberg, 1976). Future work will clarify the contexts within which MSTd cells respond to local object or global motion and how this may influence navigation.

The model of MT⁺/MSTd that we present in this article demonstrates that human heading biases estimated for approaching and non-approaching IMOs can be explained using motion pooling and template matching in a competitive network while remaining consistent with known neurophysiology. Differential motion processing is not necessary to explain these data in the presence of IMOs.

Acknowledgments

The authors thank Chris Pack and Bill Warren for helpful discussions and two anonymous reviewers for useful feedback. OWL and EM were supported in part by CELEST, an NSF Science of Learning Center (NSF SBE-0354378 and NSF OMA-0835976). OWL, EM, and NAB were supported in part by the Office of Naval Research (ONR N00014-11-1-0535).

Commercial relationships: none.

Corresponding author: N. Andrew Browning.

Email: buk@bu.edu.

Address: 677 Beacon st., Boston, MA 02215, USA.

Footnotes

¹ Assuming a more typical 1280 × 1024 monitor with a 5:4 physical aspect ratio did not impact our results.

²We assume that the objects of Royden and Hildreth (1996) had a constant size in pixels in a particular trial frame, and the object size in degrees was estimated at the center of the screen.

References

- Baloch, A. A., & Grossberg, S. (1997). A neural model of high-level motion processing: Line motion and form-motion dynamics. *Vision Research*, 37, 3037–3059. [PubMed]
- Beintema, J. A., & van den Berg, A. V. (1998). Heading detection using motion templates and eye velocity gain fields. *Vision Research*, 38, 2155–2179. [PubMed]
- Berezovskii, V. K., & Born, R. T. (2000). Specificity of projections from wide-field and local motion-processing regions within the middle temporal visual area of the owl monkey. *Journal of Neuroscience*, 20, 1157–1169. [PubMed]
- Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*, 28, 157–189. [PubMed]
- Boussaoud, D., Ungerleider, L. G., & Desimone, R. (1990). Pathways for motion analysis: Cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. *Journal of Comparative Neurology*, 296, 462–495. [PubMed]
- Britten, K. H. (2008). Mechanisms of self-motion perception. *Annual Review of Neuroscience*, 31, 389–410. [PubMed]
- Browning, N. A., Grossberg, S., & Mingolla, E. (2009a). Cortical dynamics of navigation and steering in natural scenes: Motion-based object segmentation, heading, and obstacle avoidance. *Neural Networks*, 22, 1383–1398. [PubMed]
- Browning, N. A., Grossberg, S., & Mingolla, E. (2009b). A neural model of how the brain computes heading from optic flow in realistic scenes. *Cognitive Psychology*, 59, 320–356.
- Chey, J., Grossberg, S., & Mingolla, E. (1997). Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. *Journal of the Optical Society of America A*, 14, 2570–2594.
- Chey, J., Grossberg, S., & Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research*, 38, 2769–2786. [PubMed]
- Churchland, M. M., Priebe, N. J., Lisberger, S. G., Priebe, M. M. N. J., & Lis, S. G. (2005). Comparison of the spatial limits on direction selectivity in visual areas MT and V1. *Journal of Neurophysiology*, 93, 1235–1245.
- Duffy, C. J., & Wurtz, R. H. (1991a). Sensitivity of MST neurons to optic flow stimuli: I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, 65, 1329–1345. [PubMed]
- Duffy, C. J., & Wurtz, R. H. (1991b). Sensitivity of MST neurons to optic flow stimuli: II. Mechanisms of response selectivity revealed by small-field stimuli. *Journal of Neurophysiology*, 65, 1346–1359. [PubMed]
- Duffy, C. J., & Wurtz, R. H. (1993). An illusory transformation of optic flow fields. *Vision Research*, 33, 1481–1490. [PubMed]
- Duffy, C. J., & Wurtz, R. H. (1995). Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. *Journal of Neuroscience*, 15, 5192–5208. [PubMed]
- Duijnhouwer, J., Beintema, J. A., van den Berg, A. V., & Wezel, R. J. A. van. (2006). An illusory transformation of optic flow fields without local motion interactions. *Vision Research*, 46, 439–443. [PubMed]
- Duijnhouwer, J., Van Wezel, R. J. A., & van den Berg, A. V. (2008). The role of motion capture in an illusory transformation of optic flow fields. *Journal of Vision*, 8(4):27, 1–18, <http://www.journalofvision.org/content/8/4/27>, doi:10.1167/8.4.27. [PubMed] [Article]
- Eifuku, S., & Wurtz, R. H. (1998). Response to motion in extrastriate area MSTl: Center–surround interactions. *Journal of Neurophysiology*, 80, 282–296. [PubMed]
- Elder, D. M., Grossberg, S., & Mingolla, E. (2009). A neural model of visually guided steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1501–1531. [PubMed]
- Fajen, B. R., & Kim, N.-g. (2002). Perceiving curvilinear heading in the presence of moving objects. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 1100–1119.
- Gattass, R., & Gross, C. G. (1981). Visual topography of striate projection zone (MT) in posterior superior temporal sulcus of the macaque. *Journal of Neurophysiology*, 46, 621–638. [PubMed]
- Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233, 1416–1419.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum.
- Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, L11, 213–258.

- Grossberg, S., Mingolla, E., & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, 41, 2521–2553.
- Gu, Y., Fetsch, C. R., Adeyemo, B., DeAngelis, G. C., & Angelaki, D. E. (2010). Decoding of MSTd population activity accounts for variations in the precision of heading perception. *Neuron*, 66, 596–609. [PubMed]
- Hatsopoulos, N., & Warren, W. (1991). Visual navigation with a neural network. *Neural Networks*, 4, 303–317. [Article]
- Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9, 181–197. [PubMed]
- Heeger, D. J., & Jepson, A. (1990). Visual perception of three-dimensional motion. *Neural Computation*, 2, 129–137. [Article]
- Hildreth, E. C. (1992). Recovering heading for visually-guided navigation. *Vision Research*, 32, 1177–1192.
- Lappe, M., & Krekelberg, B. (1998). The position of moving objects. *Perception*, 27, 1437–1449. [PubMed] [Article]
- Lappe, M., & Rauschecker, J. P. (1993). A neural network for the processing of optic flow from ego-motion in man and higher mammals. *Neural Computation*, 5, 374–391. [Article]
- Levine, D. S., & Grossberg, S. (1976). Visual illusions in neural networks: Line neutralization, tilt after effect, and angle expansion. *Journal of Theoretical Biology*, 61, 477–504. [PubMed]
- Logan, D. J., & Duffy, C. J. (2006). Cortical area MSTd combines visual cues to represent 3-D self-movement. *Cerebral Cortex*, 16, 1494–1507. [PubMed]
- Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B: Containing Papers of a Biological Character. Royal Society (Great Britain)*, 208, 385–397. [PubMed]
- Maunsell, J. H., & van Essen, D. C. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience*, 3, 2563–2586. [PubMed]
- Mineault, P. J., Khawaja, F. A., Butts, D. A., & Pack, C. C. (2012). Hierarchical processing of complex motion along the primate dorsal visual pathway. *Proceedings of the National Academy of Sciences of the USA*, in press.
- Nassi, J. J., & Callaway, E. M. (2006). Multiple circuits relaying primate parallel visual pathways to the middle temporal area. *Journal of Neuroscience*, 26, 12789–12798. [Abstract]
- Nelissen, K., Vanduffel, W., & Orban, G. A. (2006). Charting the lower superior temporal region, a new motion-sensitive region in monkey superior temporal sulcus. *Journal of Neuroscience*, 26, 5929–5947. [PubMed]
- Orban, G. (2008). Higher order visual processing in macaque extrastriate cortex. *Physiological Reviews*, 88, 59–89.
- Pack, C., Grossberg, S., & Mingolla, E. (2001). A neural model of smooth pursuit control and motion perception by cortical area MST. *Journal of Cognitive Neuroscience*, 13, 102–120. [PubMed]
- Pack, C., & Mingolla, E. (1998). Global induced motion and visual stability in an optic flow illusion. *Vision Research*, 38, 3083–3093. [PubMed]
- Perrone, J. A., & Krauzlis, R. J. (2008). Vector subtraction using visual and extraretinal motion signals: A new look at efference copy and corollary discharge theories. *Journal of Vision*, 8(14):24, 1–14, <http://www.journalofvision.org/content/8/14/24>, doi:10.1167/8.14.24. [PubMed] [Article]
- Perrone, J. A., & Stone, L. S. (1994). A model of self-motion estimation within primate extrastriate visual cortex. *Vision Research*, 34, 2917–2938. [PubMed]
- Raudies, F., & Neumann, H. (accepted for publication). A review and evaluation of methods estimating ego-motion. *Computer Vision and Image Understanding*.
- Rieger, J. H., & Lawton, D. T. (1985). Processing differential image motion. *Journal of the Optical Society of America A, Optics and Image Science*, 2, 354–360. [PubMed]
- Royden, C. S. (1997). Mathematical analysis of motion-opponent mechanisms used in the determination of heading and depth. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 14, 2128–2143. [PubMed]
- Royden, C. S. (2002). Computing heading in the presence of moving objects: A model that uses motion-opponent operators. *Vision Research*, 42, 3043–3058. [PubMed]
- Royden, C. S. (2004). Modeling observer and object motion perception. In L. M. Vaina, S. A. Beardsley, & S. K. Rushton (Eds.), *Optic flow and beyond* (pp. 131–153). Dordrecht: Kluwer.
- Royden, C. S., & Conti, D. (2003). A model using MT-like motion-opponent operators explains an illusory transformation in the optic flow field. *Vision Research*, 43, 2811–2826. [Article]
- Royden, C. S., Crowell, J. A., & Banks, M. S. (1994). Estimating heading during eye movements. *Vision Research*, 34, 3197–3214.

- Royden, C. S., & Hildreth, E. C. (1996). Human heading judgments in the presence of moving objects. *Perception & Psychophysics*, 58, 836–856.
- Royden, C. S., & Hildreth, E. C. (1999). Differential effects of shared attention on perception of heading and 3-D object motion. *Perception & Psychophysics*, 61, 120–133. [PubMed]
- Sato, N., Kishore, S., Page, W. K., & Duffy, C. J. (2010). Cortical neurons combine visual cues about self-movement. *Experimental Brain Research*, 206, 283–297. [Abstract]
- Saunders, J. A., & Niehorster, D. C. (2010). A Bayesian model for estimating observer translation and rotation from optic flow and extra-retinal input. *Journal of Vision*, 10(10):7, 1–22, <http://www.journalofvision.org/content/10/10/7>, doi:10.1167/10.10.7. [PubMed] [Article]
- Sincich, L. C., & Horton, J. C. (2005). The circuitry of V1 and V2: Integration of color, form, and motion. *Annual Review of Neuroscience*, 28, 303–326. [PubMed]
- Warren, W. H. (1998). The state of flow. In T. Wanatabe (Ed.), *High-level motion processing* (pp. 315–358). Cambridge, MA: MIT Press.
- Warren, W. H. (2009). How do animals get about by vision? Visually controlled locomotion and orientation after 50 years. *British Journal of Psychology*, 100, 277–281. [PubMed]
- Warren, W. H., Blackwell, A. W., Kurtz, K. J., Hatsopoulos, N. G., & Kalish, M. L. (1991). On the sufficiency of the velocity field for perception of heading. *Biological Cybernetics*, 65, 311–320.
- Warren, W. H., & Hannon, D. J. (1990). Eye movements and optical flow. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 7, 160–169. [PubMed]
- Warren, W. H., Kay, B. A., Zosh, W. D., Duchon, A. P., & Sahuc, S. (2001). Optic flow is used to control human walking. *Nature Neuroscience*, 4, 213–216. [PubMed]
- Warren, W. H., & Kurtz, K. J. (1992). The role of central and peripheral vision in perceiving the direction of self-motion. *Perception & Psychophysics*, 51, 443–454. [PubMed]
- Warren, W. H., Morris, M. W., & Kalish, M. (1988). Perception of translational heading from optical flow. *Journal of Experimental Psychology*, 14, 646–660. [PubMed]
- Warren, W. H., & Saunders, J. A. (1995). Perceiving heading in the presence of moving objects. *Perception*, 24, 315–331. [PubMed]
- Xiao, D. K., Marcar, V. L., Raiguel, S. E., & Orban, G. A. (1997). Selectivity of macaque MT/V5 neurons for surface orientation in depth specified by motion. *The European Journal of Neuroscience*, 9, 956–964. [PubMed]
- Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, 20, 6594–6611. [PubMed]