Reviewed Preprint
v1 • April 16, 2026
Not revised

Reviewed Preprint
v2 • August 12, 2026
Revised by authors

For correspondence:
j.lopezmoliner@ub.edu

Competing interests: No competing interests declared

Funding: See page 19

Reviewing editor: Marisa Carrasco, New York University, United States

Beyond the Focus of Expansion: Retinal curl as a functional signal for heading estimation

Kontessa I Zorpala, Joan López-Moliner

Institute of Neurosciences, Universitat de Barcelona, Barcelona, Spain

eLife Assessment

This study provides an **important** and biologically plausible account of how human perceptual judgments of heading direction are influenced by a specific pattern of motion in optic flow fields known as retinal curl. By combining psychophysical experiments and neural modeling, the authors demonstrate that what was previously considered an incidental, 'nuisance' signal actually serves as a functional control signal for estimating heading and steering toward a fixated target. The evidence supporting the role of curl signals is **convincing** and advances our understanding of vision-based navigation at the behavioral level. A particular strength of the work is the direct manipulation of curl within flow fields, demonstrating that it effectively cancels or reverses heading biases. This work provides an invaluable framework for future exploration into the neural mechanisms of steering control based on retinal curl.

https://doi.org/10.7554/eLife.110770.2.sa4

Abstract

Prevailing models aiming at explaining heading assume that humans need to recover the Focus of Expansion (FoE) while accounting for eye-movement-induced rotation. We propose an alternative: the visual system utilizes mean retinal curl from fixations as a surrogate signal for heading, rendering the explicit recovery of the FoE unnecessary. Stationary participants viewed simulated walking paths on a large screen while fixating on points on the projected ground texture at varying eccentricities —a natural behavior inducing sustained retinal curl. Participants continuously reported perceived heading in 3D scene coordinates. To isolate the role of retinal curl, we employed a real-time manipulation that kept translational flow constant while the foveal curl component was either unaltered, cancelled, or overcancelled. Under natural conditions (unaltered), participants exhibited systematic heading biases opposite the direction of gaze. Crucially, these biases vanished when we cancelled the expected curl, and flipped when we overcancelled it, identifying retinal curl as the specific driver of perceptual bias. We modeled these results using a simple feedback controller and a ring-attractor neural network featuring gaze-contingent inhibition and a ‘straight-ahead’ prior. These findings suggest the brain exploits the geometry of gaze stabilization to simplify navigation, treating retinal curl as a functional signal rather than noise to be filtered.

Introduction

Humans and many vertebrates rely on stabilizing gaze on regions of interest in the world to achieve accurate perception. This fixation strategy, combined with the continuous eye, head, and body movements that accompany natural behavior, generates highly structured patterns of motion on the retina that can, in principle, be exploited to control locomotion. A particularly simple and influential case arises when we move approximately in the same direction as we are looking: in this situation the focus of expansion (FoE), the point in the optic flow from which motion vectors appear to radiate, directly specifies the heading or trajectory direction (Gibson,

1950). The idea that the visual system uses the FoE to recover heading has subsequently dominated both theoretical and empirical work, including the interpretation of neural activity in primate medial superior temporal (MST) area, where neurons exhibit FoE-like tuning (Bremmer et al., 2017; Britten, 2008; Duffy and Wurtz, 1991; Kaminiarz et al., 2014), as well as psychophysical findings showing that humans can make highly accurate heading judgements from radial optic flow (e.g Warren and Hannon (1990); Warren and Hannon (1988); Warren et al. (1988); van den Berg (1992); Li and Warren (2000)).

However, rotating the eyes or head to look at an off-path object adds a rotational component to the retinal flow. This combined motion distorts the pure expansion pattern, shifting the FoE so it no longer aligns with the true heading. Despite this distortion, observers can generally maintain accurate heading judgments during both active eye movements (Banks et al., 1996; Royden et al., 1992) and simulated rotations—where the rotational visual component is artificially rendered on a screen without the observer actually moving their eyes—provided sufficient 3D depth is visible (Cutting, 1986; Grigo and Lappe, 1999; Li and Warren, 2002; Warren and Hannon, 1988). To explain how humans distinguish where they are heading from where they are looking, classic models suggest the brain decomposes the flow or subtracts the rotation to recover the pure FoE. This de-rotation is achieved either by exploiting visual cues alone (Beintema et al., 2004; Cutting et al., 1992; Heeger and Jepson, 1992; Lappe and Rauschecker, 1993; Longuet-Higgins and Prazdny, 1980; Perrone and Stone, 1994), or by integrating extra-retinal signals (Beintema and van den Berg, 1998; Koenderink and van Doorn, 1981; Lappe, 1998; Royden et al., 1994; van den Berg and Beintema, 2000). Yet, this compensation is not perfect: participants often make systematic heading errors during these simulated rotations devoid of extra-retinal cues (Banks et al., 1996), particularly when depth is absent (such as when viewing a flat, single fronto-parallel wall) (Grigo and Lappe, 1999; Rieger and Toet, 1985) or when the tracking target moves independently of the ground plane (Royden et al., 1992).

The fact that recovering self-motion from rotational flow requires some type of parallax (e.g. Rieger and Toet (1985); Stone and Perrone (1997); Grigo and Lappe (1999)), depth order (van den Berg and Brenner, 1994) or rigid 3D structure to guide heading (Cutting, 1986; Li and Warren, 2000; Warren, 1998) and steering (Li and Warren, 2002; Wann and Swapp, 2000) directly challenges visual decomposition models. If the brain simply subtracted rotation from translation, heading could theoretically be recovered even from flat, fronto-parallel planes. The systematic errors observed when depth is absent—unless extra-retinal cues are available—suggest the brain bypasses global de-rotation. Pure visual decomposition should function regardless of 3D depth or whether the rotation stems from an active eccentric fixation.

Indeed, behavioral evidence shows that humans successfully recover heading specifically when this rotational flow arises from tracking an eccentric, grounded target (Perrone and Stone, 1994; van den Berg, 1993; Warren and Hannon, 1988). In natural navigation, this type of continuous gaze stabilization is ubiquitous (Matthis et al., 2018). Instead of creating a nuisance artifact that must be subtracted, this gaze behavior generates a useful, depth-dependent rotational component in the form of structured spiral gradients—the retinal curl. In real-world locomotion, foveal curl provides a more reliable gaze-relative signal than the highly variable, head-centered FoE (Matthis et al., 2022). Crucially, this curl extends beyond estimating an instantaneous heading to specify a "future path" (Wann and Swapp, 2000; see Li, 2025 for a recent review), offering predictive information that is highly advantageous for active steering. The use of curl aligns with reinterpreting neurophysiological evidence: rather than utilizing segregated expansion and rotation channels, MSTd neurons would be tuned to a continuum of spiral motion patterns (Graziano et al., 1994; Layton and Browning, 2014). Such tuning suggests the visual system actively exploits the rotational gradients inherent to active fixation strategies (Angelaki and Hess, 2005; Calow and Lappe, 2008; Glennerster et al., 2001).

Here, we provide empirical evidence that the visual system utilizes retinal curl as a primary control variable for locomotor navigation. Participants reported their perceived trajectory in world-centered (3D scene) coordinates during simulated walking through a naturalistic optic flow field (see Figure 1A). This paradigm revealed a robust, systematic bias directed strictly opposite

to the direction of gaze: fixating an eccentric ground target to the left induces a rightward shift in perceived heading, and vice versa. This effect, which requires sustained exposure and differs qualitatively from classic simulated-rotation errors (Supplementary Videos S1 and S2), is precisely predicted by gaze-centered flow geometry. Crucially, experimentally cancelling the retinal curl eliminated the bias entirely, confirming its causal role. To mechanistically ground these findings, we show that a simple feedback controller driven by mean image curl faithfully reproduces human steering behavior across diverse path conditions. Finally, we provide a neurally plausible implementation illustrating how parietal circuits might transform local foveal curl into global heading estimates via recurrent dynamics and the competitive integration of sensory evidence and spatial priors. Consequently, we propose a redefined role for extra-retinal information. Rather than simply "de-rotating" the flow field, eye position signals serve to scale and interpret retinal curl. The visual system would exploit the magnitude of foveal curl directly, transforming this structured spiral gradient into calibrated steering commands without uncovering a hidden FoE.

Results

Observers viewed simulated forward locomotion along different paths over a ground plane (Figure 1B). During each trial, they maintained continuous fixation on a specific ground-embedded target, which was either aligned with their path (straight ahead) or laterally displaced (2 m or 4 m to the left or right); generating different mean curl distributions (see Figure 1C-E). While keeping this stabilized gaze, participants continuously reported their perceived heading in world coordinates, allowing us to reconstruct their full perceived trajectory over time.

We first assessed the quality of fixation to ensure gaze stability. Across subjects, the median trial-averaged deviation from the fixation target ranged from 1.02° to 1.55°. Gaze-to-target distance exceeded 3° in only 0.46% of trials; these were excluded from further analysis. In addition to spatial accuracy, we calculated the retinal slip (the velocity of the target image on the retina). Median horizontal slip ranged from 0.12 to 0.27°/*s* and vertical slip from 0.12 to 0.23°/*s*, confirming that gaze was effectively stabilized throughout the simulated locomotion.

Figure 2 shows the instantaneous perceived heading across different gaze and path conditions where retinal flow was unaltered. Paths were reconstructed in world (3D scene) coordinates by integrating the raw responses (see *Computation of Perceived Path* in the Methods section and Figure 2 Suppl. Figure 1 for an example of raw responses). The thin traces show this pattern for individual observers, and the thick coloured traces show the group means. Figure 2 illustrates a clear and systematic lateral bias in perceived heading as a function of the inital gaze direction. When observer-simulated motion is straight but fixated an eccentric point on the ground (central vertical panels, blue and green lines), their reported instantaneous heading consistently shifted away from the fixation point: leftward fixation led to rightward heading reports, and rightward fixation produced the opposite pattern. The perceived final lateral displacement is significantly different from 0 (straight ahead, $t(11)=5.172$, $p<0.001$) and the magnitude of the bias was not different between sides ($F(1,35)<1$, $p=0.4$). This effect was present both when gaze was displaced by 2 m (top row) and 4 m (bottom row), and the effect scales with the lateral distance of the fixation point from the trajectory path (the final lateral distance between 2 m and 4 m condition was marginally significant, $F(1,35)=3.49$, $p=0.07$). As we shall see below, the magnitude of the bias is well accounted for by the present mean curl.

Note that when both simulated motion and gaze are straight ahead (orange traces), there is essentially no bias: path is perceived straight. This is consistent with curl staying near zero throughout the trial (see Figure 1D). In this case, heading and gaze remain aligned and perceived trajectory stays close to the physical straight path.

Although the bias is more easily interpretable in the straight-ahead condition, systematic effects due to gaze direction are also present in curved paths. Perceived curvature was systematically *underestimated* when gaze was directed into the direction of the physical curve: for leftward curves, this underestimation is most evident when gaze is to the left (blue traces), and for rightward curves, when gaze is to the right (green traces).

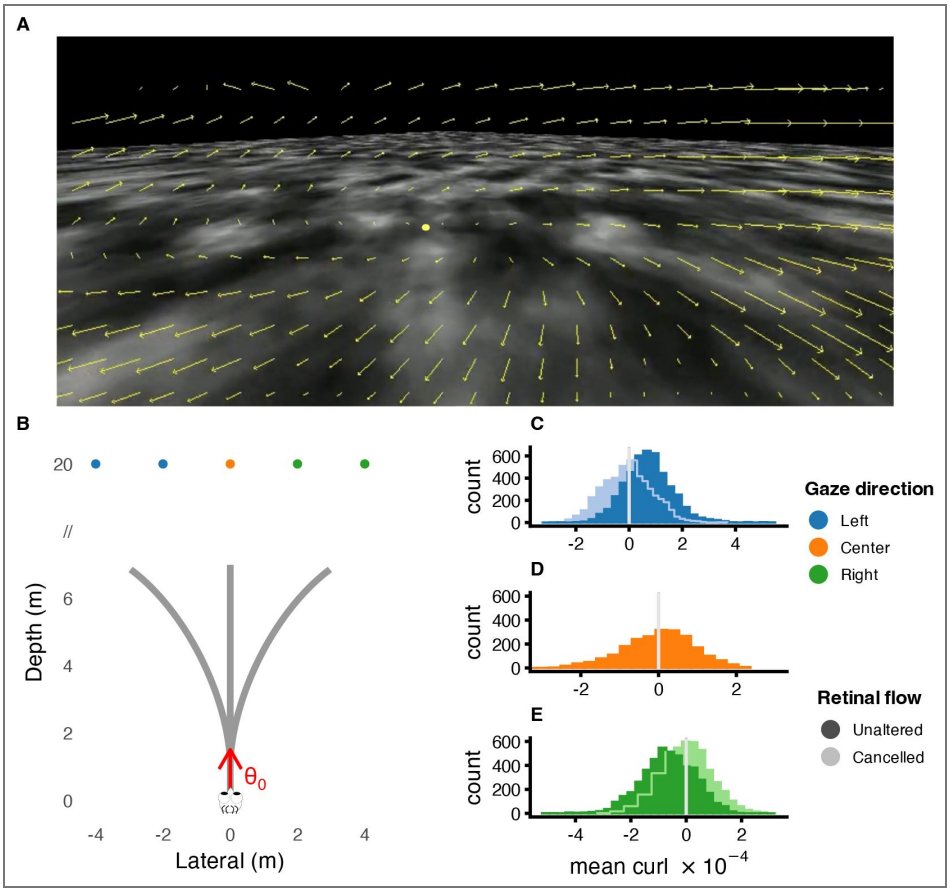


Figure 1. Ground texture, trajectories and retinal curl distributions across conditions.

(A) Snapshot of the ground texture based on simplex noise. Yellow lines indicate optic flow vectors computed using the Farnebäck algorithm. A clear rotational component (curl) is visible, consistent with the observer looking at a point (yellow dot) located to the left of the simulated path. The Focus of Expansion is shifted in the direction of the gaze. (B) Schematic of the experimental trajectories. Position (0, 0) represents the starting point of simulated locomotion. The red arrow indicates the initial heading (θ_0), which was aligned longitudinally with the 3D scene (world coordinates). Participants were instructed to report their perceived heading within this world-centered reference frame. The five colored dots mark the fixation points in world coordinates at the beginning of each trial (20 m ahead of the observer). (C–E) Distributions of mean retinal curl across trials for left-, center-, and right-gaze conditions, respectively. Dark filled bars indicate the unaltered curl condition, while lighter bars and outlined steps represent the cancelled curl condition. The vertical grey line denotes zero curl.

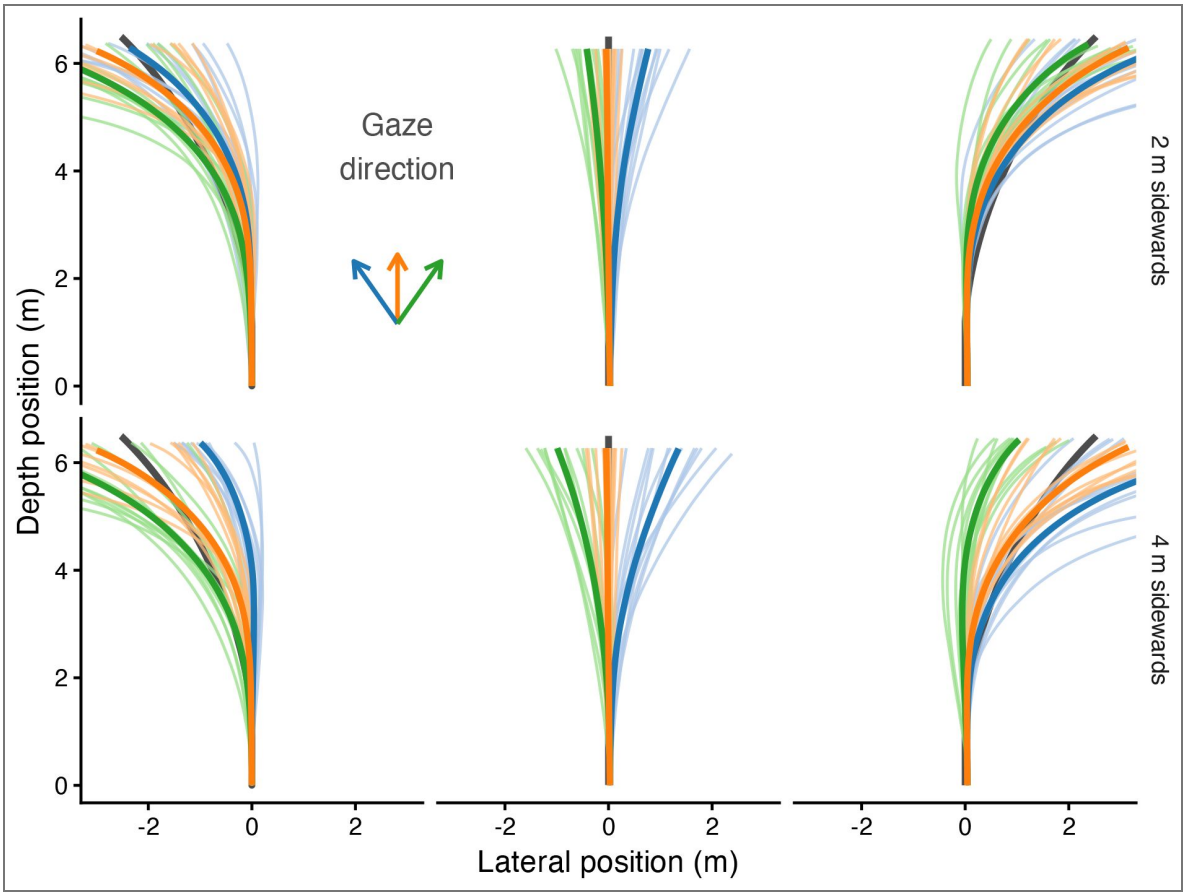


Figure 2. **Perceived heading**

Reported instant directions are plotted when fixating eccentric points on the ground. Rows: fixation 2 m (top) and 4 m (bottom) to the side. Columns: different physical path conditions (centre column = straight path). Colours code for initial gaze direction (left / centre / right). Thick coloured lines denote the mean across observers. Thin coloured lines denote individual observers. Dark grey denotes physical paths. Axes show lateral position (x-axis) versus depth position (y-axis)

This follows directly from the geometry of gaze-stabilization. As the observer proceeds along a curved path while looking toward the inside of the turn, the heading (θ_t) and gaze directions (ψ_t) become progressively aligned. This alignment causes the magnitude of retinal curl to naturally diminish over time. In our proposed controller (see Methods: control model), the visual system treats this curl as an error signal ($\Phi_t = \psi_t - \theta_t$) that drives the required turning rate: the yaw command to change heading (θ_t) scales with this error ($\theta_t = K_p\phi_t$, see Equation 6). Therefore, as the curl signal 'dies out' due to the alignment of gaze and heading, the model predicts a corresponding reduction in the perceived turning command—resulting in a progressive flattening of the perceived path. This is exactly the pattern observed empirically, and it is symmetric for left and right curves. By contrast, when gaze is directed opposite to the direction of curvature, the gaze and heading do not align —Φ does not collapse toward zero—, the curl remains sustained, and no underestimation of curvature is expected or observed.

Effects of flow manipulation

To establish a direct causal link between retinal curl and the observed heading biases, we introduced specific manipulations to the optic flow field in interleaved trials. By artificially cancelling or over-cancelling the natural rotational flow generated during eccentric fixation, we tested whether the participants' perceived trajectories would predictably flatten or reverse in agreement with the exposed curl.

Figure 3 shows the reported instant heading for the different flow manipulations. We also show the unaltered condition (green) again for the sake of comparison. Importantly, when curl is cancelled (cyan), these biases essentially disappear – the perceived path remains much closer to the physical trajectory across gaze directions. Note that the central row (straight-ahead simulated motion) provides the clearest demonstration: the unaltered curl shows the previously reported lateral biases, but when curl is cancelled (counteracted) the perceived trajectory becomes near-veridical. This confirms that the mean curl signal itself – not gaze eccentricity per se – is the functional driver of the bias. In the over-cancelled condition (purple), the pattern reverses, with biases in the opposite direction. To maintain the full factorial design (Eccentricity × Flow Manipulation × Heading), we included trials where gaze was centered and straight ahead. Although no natural curl is generated at this zero-eccentricity position, we purposely introduced curl at rotation speeds corresponding to the 'cancelled' and 'over-cancelled' conditions (simulating random leftward or rightward gaze). As shown in the central panel of Figure 3, this introduced flow induced the predicted biases despite the absence of gaze eccentricity, confirming that the retinal flow pattern, rather than the physical orientation of the eyes, is the primary driver of the perceived heading shift.

The last column in Figure 3 shows the average 2D displacement (in meters) required for the observed paths to align with the physical ones. This measure gives an idea of the similarity between the reported paths and the physical exposed trajectories in the experiment. In general, the difference between perceived and physical paths is smaller when the path is straight (central panel of last column in Figure 3). Consistent with the reported paths in the different flow manipulation conditions, the displacement is smaller in the cancelled flow curl (color-coded) for all headings (different rows). This resulted in a significant quadratic effect of Flow manipulation, $\beta = 0.09$, SE = 0.023, $t(526) = 3.997$, $p < .001$, indicating that mean deviation was lower in the cancelled condition compared with the unaltered and overcancelled conditions.

Fitting the controller

To assess how well the curl-based controller accounts for reported perceived headings, we fitted the model as described in the Methods. The fit incorporates the mean image curl computed via Equation 1 from the experimental videos, assuming perfect gaze stabilization on the fixation point. Fits were performed for both the aggregate data and individual participants (see Figure 4 and its corresponding supplemental figures in the SI). As shown in Figure 4, the model with separate parameters per condition captures the average perceived trajectories (thin lines) with remarkable accuracy (thick red lines). In these separate fits, the mean 2D deviation per step was

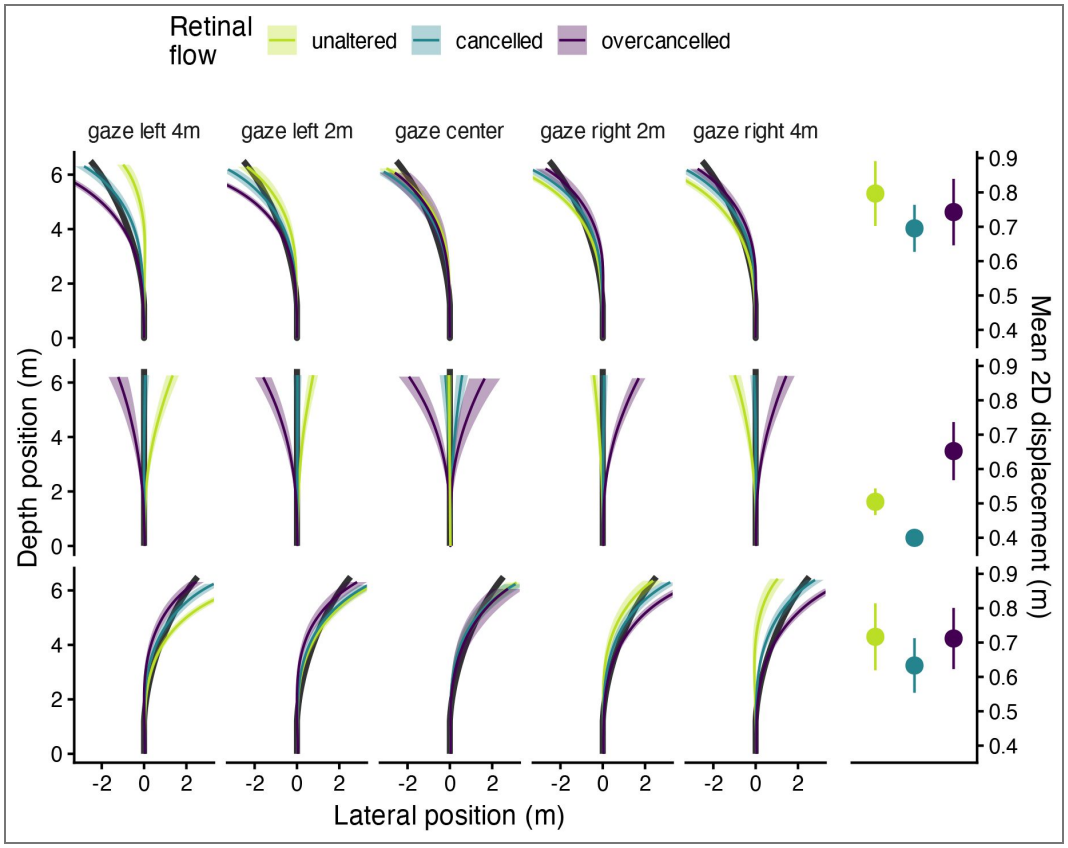


Figure 3. Average perceived heading for each gaze condition (columns) and each physical path curvature (rows), separately for the three retinal–flow manipulation conditions: unaltered curl (green), cancelled curl (cyan), and over-cancelled curl (purple).

Shaded envelopes indicate between-observer variability (95%-CI). Note the heading biases in the gaze center condition/middle row. Unexpected positive or negative curl was added in some trials to mantain a full factorial design (see main text for more details). The right axis applies to the last column and illustrates the mean 2D displacement between observed and physical paths. This measurement indicates the mean displacement that is required for the observed path to align with the physical one. The displacement is shown for the different flow manipulations (color-coded). The error bars indicate between-observer variability (95%-CI).

minimal, ranging from 0.066 m in the cancelled flow condition to 0.080 m in the unaltered condition (see Table 1 in Appendix 3). Statistical analysis via ANOVA on the individual fits confirmed that while separate fits yielded a significantly smaller loss than the join approach (F (1, 55) = 206.27, p < 0.001), flow manipulation itself had only a marginal effect on fit quality (F (2, 55) = 2.69, p = 0.077). Interestingly, the cancelled condition showed a slightly smaller loss than the other two (Quadratic contrast t(55) = 2.356, p = 0.022).

We also evaluated a joint fit—using only two parameters to account for all heading and eccentricity conditions. We view this joint fit not as a perfect model, but as a deliberate limit test of the curl-based model. In reality, the control "Gain" (the weighting of the curl signal) is likely tuned dynamically based on task demands and sensory confidence. By forcing a single, fixed linear relationship between retinal curl and steering effort, we intentionally ignore this contextual variability to determine how much behavior can be explained by a minimalist controller. Although this compromise naturally results in a larger 2D deviation, it successfully captures the core qualitative trends: a) tracking near-straight trajectories when curl is cancelled; b) reproducing the systematic underestimation of curvature when gaze aligns with the curve; and c) predicting the reversal of bias directions in the over-cancelled condition (purple lines). Upon closer analysis, the most pronounced discrepancies in the joint model occur in this 'over-cancelled' condition, where sensory flow evidence becomes ecologically inconsistent with extra-retinal gaze direction, likely requiring complex sensory re-weighting. We therefore frame this joint model, not as a perfect model, but as a parsimonious, first-order approximation. Despite the higher deviation inherent in fixing parameters, its statistically superior (lower) AIC measures (Table 1 in Appendix 3) demonstrate that a minimal, unified curl-based mechanism provides a fairly good approximation for heading perception across diverse conditions.

Neural simulations

Having established that a curl-based controller successfully accounts for heading responses, we next sought to determine how this mechanism might be implemented in a neurophysiologically plausible way. To do so, we developed a recurrent ring network model designed to transform local foveal flow into global heading estimates (see appendix 2). This architecture explicitly incorporates the role of active gaze by modeling the foveal curl as a localized inhibitory drive on the network's activity. Because our primary aim with this neural model is to mechanistically explain the origin of the systematic opposite-gaze bias, we focused our simulations exclusively on the straight-path conditions. The model processed optic flow directly from the experimental video sequences—assuming perfect gaze stabilization—and simulated the continuous neural dynamics from local motion extraction to the linear decoding of the population vector.

To identify the specific drivers of the heading bias, we systematically explored the network's parameter space. We varied the prior strength ($I_0 \in [0.03, 0.24]$), the prior width ($\sigma_p \in [0.18, 0.25]$), the spatial extent of gaze inhibition ($\sigma_k \in [0.02, 0.04, 0.08, 0.12]$) and the inhibitory gain ($K_p \in [0.4, 0.8]$). Table 1 in the appendix 2 shows the range for the different parameters used in the simulations.

Simulation results

The network dynamics provide a clear mechanistic explanation for the systematic heading bias observed during eccentric gaze. In Figure 5, we observe the network components for a trial in which the gaze direction is 4 m to the left during forward translation. Figure 5A shows the recurrent connectivity following a standard Mexican hat profile, characterized by local excitation and broader surround inhibition. The activity bump does not settle at the objective straight-ahead (0°) but stabilizes at a positive (rightward) heading offset, as illustrated in Figure 5B. This displacement is driven by the competition shown in the mechanism panel (Figure 5C): the gaze-centered inhibitory drive (red dashed line) suppresses the leftward portion of the ring, effectively 'pushing' the neural activity bump (black line) away from the gaze location and to the right of the straight-ahead prior (blue dotted line). The Phase Portrait (Figure 5 D) confirms this as a stable state; the drift dynamics $\left(\frac{d\theta}{dt}\right)$ converge toward a stable fixed point where the curve crosses zero at

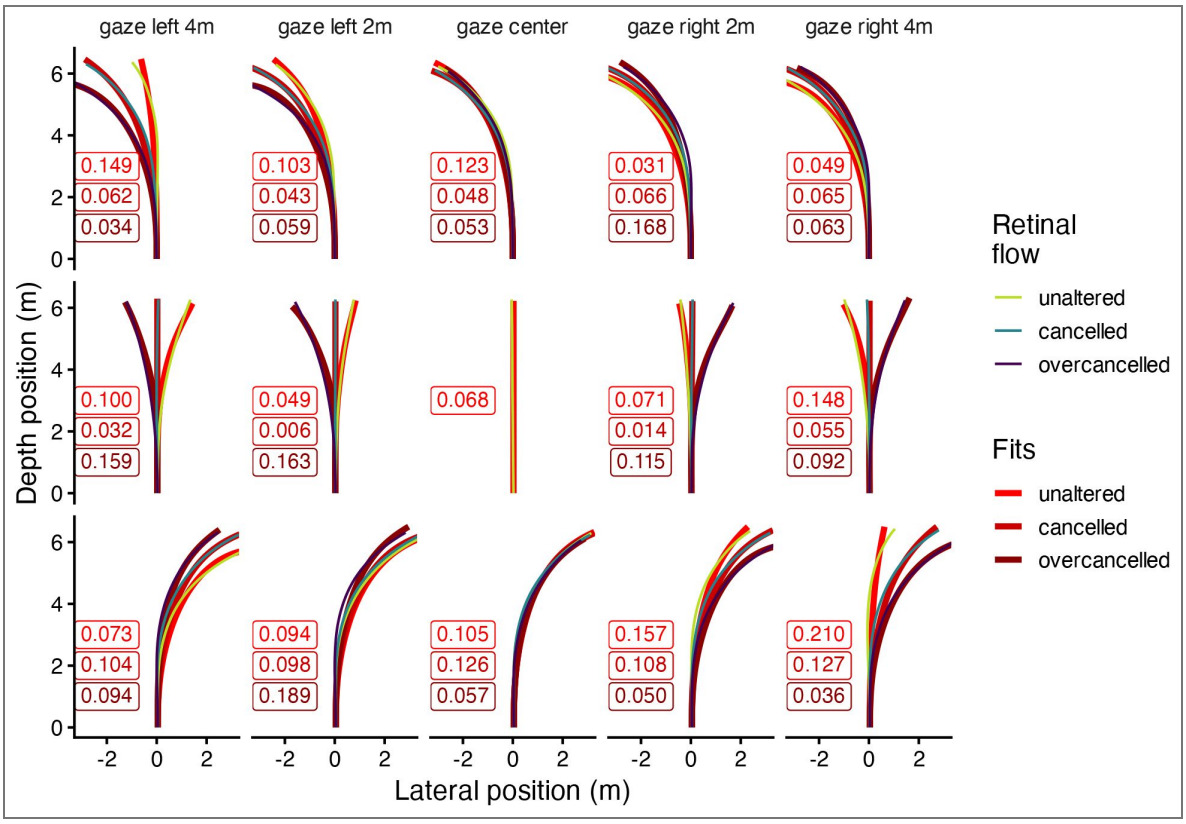


Figure 4. Separate fits of the controller.

Average perceived heading across participants for each gaze condition (columns) and each physical path curvature (rows), separately for the three retinal–flow manipulation conditions (thiner solid lines): unaltered curl (green), cancelled curl (cyan), and over-cancelled curl (purple). The different red thicker solid lines denote the best fit of the controller. The numbers in each panel indicate the average lateral deviation per step between the fit and the observed heading. Note that for the centered gaze and straight path, only the unaltered flow condition was fitted.

a rightward offset. Together, these results demonstrate how localized inhibition, within a parietal ring attractor, transforms gaze-contingent sensory signals into a shifted global heading estimate. For results across a range of eccentricities during forward translation, see Figure 5 Suppl. Figure 1 in the SI. The presence of temporal oscillations in the activity bump when gaze is aligned with the direction of travel (Figure 5 Suppl. Figure 1 B) or at low eccentricities (A and C) represents a state of high competition within the attractor. As seen in Figure 5 Suppl. Figure 1 B, when the gaze-centered inhibition —the sensory *pull*— overlaps directly with the straight-ahead prior —the internal *push*— it creates a central suppression that counteracts the recurrent excitation of the network. This conflict results in a concave activity profile and rhythmic fluctuations in bump intensity as the network continuously attempts to re-stabilize the representation against the central inhibitory zone. Notably, these oscillations persist when gaze is less eccentric (e.g., at -2 and 2 m in rows A and C respectively), as the proximity of the inhibitory signal to the prior creates a similar destabilizing interaction. Despite these internal dynamics, the decoded heading remains accurate due to the symmetry of the competing forces, suggesting that while representation coherence may decrease, the population vector readout remains robust to central sensory-prior conflict.

From the neural model heading estimates (θ, in pixels) we reconstructed the 3D paths using a pinhole camera model (see Heading Estimation and 3D Path Reconstruction section in SI) that are shown in Figure S17 of the SI. This figure demonstrates that the observed instant headings pattern shown in Figure 2 is best captured by a network configuration featuring a weak straight-ahead prior ($I_0 \approx 0.03$) and moderate inhibitory gains (K_p between 0.4 and 0.8). Within this parameter space, the localized, gaze-contingent inhibition effectively shifts the population activity bump, reproducing the characteristic lateral spread seen in the behavioral data. A low prior value (I_0 = 0.03 in the simulation grid) is essential. Higher prior strengths (e.g., I_0 = 0.24) overly anchor the heading estimate to the center (0 m), failing to produce the significant lateral spread observed in the human data. K_p = 0.4 closely matches the tighter trajectory fan seen in the top central panel of the empirical data, while K_p = 0.8 captures the wider lateral deviations observed in the bottom central panel, where the bias is more pronounced. Concerning the inhibition width (σ_k), values between 0.08 and 0.12 provide the most realistic structural match. Very low values (σ_k = 0.02, 0.04) produce negligible bias even at high gain, as the localized inhibition does not sufficiently overlap with the heading activity bump to drive a shift. The prior width value of σ_p = 0.18 appears to provide a more sensitive range for capturing the gaze-contingent shift compared to the wider σ_p = 0.25. These results suggest that the interplay between sensory-driven inhibition and a stabilizing straight-ahead prior within a standard Mexican-hat recurrent architecture is sufficient to account for the gaze-contingent biases observed in human heading perception.

Discussion

Heading Bias Reveals Retinal Curl as a Control Variable

We provide evidence that mean retinal curl plays a functional role in heading perception, as manipulating retinal flow predictably alters perceived trajectories. This finding suggests that curl could be used as a control signal rather than being a “nuisance” component of the optic flow that must be filtered out. While previous work has emphasized the decomposition of flow into translational and rotational components to recover heading (Beintema et al., 2004; Heeger and Jepson, 1992; Longuet-Higgins and Prazdny, 1980), we show that the visual system exploits the curl generated by gaze stabilization to perceive heading, as the sustained presence of this image curl predictably biases heading judgments. When the naturally occurring curl was counteracted, the systematic heading bias disappeared. This confirms that the bias is not an artifact of gaze eccentricity itself, but a direct consequence of the underlying flow geometry due to sustained gaze stabilization. In addition, the flow manipulation reveals a direct link between the amount of curl and perceived heading: not only did the bias opposite to fixation disappear when the curl was cancelled but also the bias re-appeared towards fixation when the curl was over-

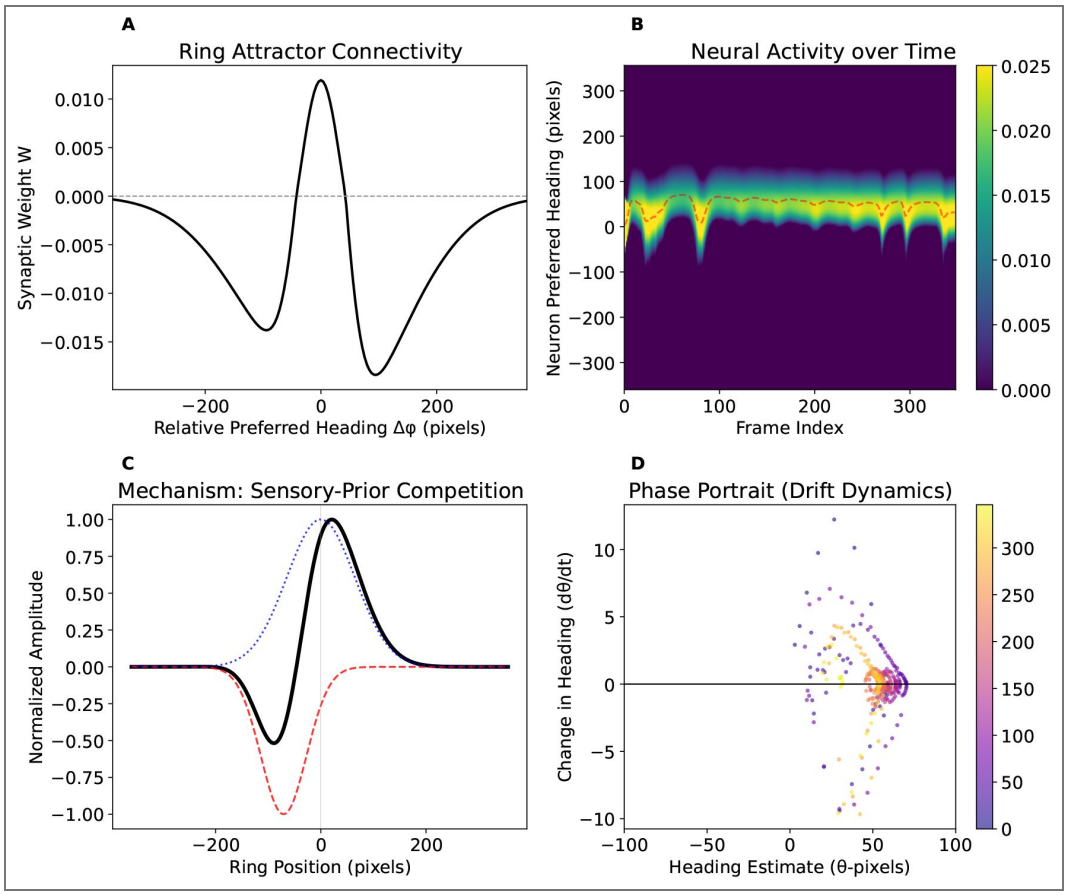


Figure 5. Neural network model of gaze-contingent heading bias.

The parameters used to produce these panels are: I_0 = 0.03, K_p = 0.4, σ_p = 0.18, σ_k = 0.12. (A) Ring attractor connectivity showing synaptic weight (W) as a function of relative preferred heading ($\Delta\Phi$) in pixels, featuring a central excitatory peak and asymmetric inhibitory sourround. While we present results using asymmetric connectivity, no significant differences were observed between asymmetric and symmetric configurations. (B) Heatmap of neural activity across neuron preferred headings (y-axis) over the frame index (x-axis), with a dashed red line indicating the decoded population heading estimate for the trial in which gaze directed 4 m to the left. (C) Mechanism of sensory-prior competition plotting normalized amplitude against ring position in pixels, illustrating the spatial alignment of the straight-ahead prior (blue dotted line), gaze-centered inhibition (red dashed line), and the resulting neural activity bump (black solid line). (D) Phase portrait showing the change in heading ($d\theta/dt$) versus the heading estimate (θ) in pixels, with a horizontal line at zero marking the convergence of the trajectory toward a stable fixed point The color bar denotes frame number.

cancelled (see paths in the middle row in Figure 3). This manipulation led to counter-intuitive findings like the reported path was closer to the physical/experimental one in the cancelled condition (last column in Figure 3)

This is consistent with Matthis et al. (2022), who observed that the head-centered FoE during natural walking is too variable to guide navigation reliably. Instead, they found that the magnitude of retinal curl provides a stable, gaze-relative trajectory signal. By experimentally isolating this curl, we provide the perceptual link demonstrating its direct effect on perceived heading. Thus, curl acts as a functional surrogate; when combined with extra-retinal gaze signals, it allows the visual system to bypass the unstable FoE in favor of a robust, rotation-based control signal.

Reference frames

Heading perception has traditionally been studied using either retino-centric (Grigo and Lappe, 1999; Rieger and Toet, 1985; Stone and Perrone, 1997) or world-centered (Banks et al., 1996; Royden et al., 1992; van den Berg, 1992; van den Berg and Brenner, 1994; Warren and Hannon, 1988) reference frames. Although reporting heading in world coordinates often feels more intuitive to observers, this extrinsic frame likely reflects the specific demands of the experimental task rather than the coordinate system of the underlying sensory information. Retinal curl inherently provides a gaze-centered signal, specifying the angular offset between the observer's trajectory and their gaze axis. To fulfill the task requirement of a world-centered judgment, observers could simply anchor the known gaze direction to the 3D scene (e.g., relative to the initial longitudinal alignment, θ_0 in Figure 1 B) and apply this curl-derived offset. Thus, observers can resolve an extrinsic trajectory without ever needing to extract a world-fixed FoE.

While this gaze-relative offset explains our data, we must consider whether an alternative, egocentric reference frame could also account for the observed opposite-to-gaze bias. Typically, gaze-dependent heading judgments bias *toward* the direction of gaze (Grigo and Lappe, 1999; Royden et al., 1992). To explain an *opposite* bias using an egocentric frame or visual direction (Rushton et al., 1998; Wann and Land, 2000), one must assume that retinal curl induces an illusory body rotation. For instance, if rightward fixation is misinterpreted as rightward body rotation, the perceived straight-ahead drifts right, causing a straight physical path to be erroneously judged to the left. However, this illusory shift is highly unlikely to explain our results for two reasons. First, egocentric visual direction relies on matching discrete targets or end-points (Rushton et al., 1998), whereas our continuous, landmark-free ground plane strongly favored flow-based processing (Wilkie and Wann, 2003). Second, interpreting foveal curl as a global body rotation creates a fundamental conflict between the exposed optic flow and the expected visual consequences of actual body rotation.

A second alternative explanation for the opposite-gaze bias lies in the under-compensation of extra-retinal signals (Freeman et al., 2010). If the brain systematically underestimates the magnitude of an eccentric eye movement, the perceived gaze is located closer to the mid-line than its true physical position. Adding an accurate curl-based visual offset to this under-compensated gaze vector would push the resulting perceived heading past the actual trajectory, generating an opposite-gaze bias. However, this extra-retinal under-compensation cannot explain the bias in our central gaze condition, where participants maintained fixation directly straight ahead while structured foveal curl was artificially inoculated into the display.

Temporal integration scales

The perception of a curved path during physically straight motion is not new (Grigo and Lappe, 1999; Royden et al., 1994, 1992; Royden, 1994; van den Berg and Brenner, 1994). Typically, longer presentation times facilitate more accurate heading judgments (Longuet-Higgins and Prazdny, 1980; Royden, 1994; Stone and Perrone, 1997; Xie and Li, 2025). However, despite our substantial trial durations (>10 s), we observed a robust bias that cannot be attributed to a lack of integration time. Notably, while Grigo and Lappe (1999) found that increasing

stimulus duration (from <1 s to 3 s) increased biases toward the direction of gaze, our findings reveal a bias in the opposite direction. This difference suggests that the nature of the bias in our study and those previously reported must obey different underlying mechanisms.

A key difference is the rotation threshold. Earlier biases emerged only at simulated rotations exceeding 1°/s, with perception remaining accurate at the lower velocities used here (Warren and Hannon, 1990). Additionally, because previous studies used brief 2–3 second stimuli, they missed that the opposite-gaze bias requires 3–5 seconds to fully manifest (Figure 2). This delay aligns with optic flow stabilization times (Warren et al., 2001) and suggests the brain integrates retinal curl over several gait cycles. Such extended processing matches evidence that global motion signals (e.g., expansion and rotation) require 1–3 second integration windows (Burr and Santoro, 2001), far surpassing the ~200 ms limit of local motion.

While instantaneous flow is critical for millisecond-scale postural balance (Bardy et al., 1999; Powell et al., 2026) and "instantaneous" retino-centric heading judgments asymptote around 400 ms (Perrone and Stone, 1994; Stone and Perrone, 1997), locomotor path control relies on a much longer integration scale. Such brief windows can be insufficient to resolve the ambiguities inherent in curved paths. Indeed, recent evidence confirms that longer stimulus durations improve heading judgments by stabilizing the trajectory estimate, rather than by enhancing the flow-parsing process itself (Xie and Li, 2025). Our use of continuous reporting—unlike the binary, post-trial responses of most prior studies—allowed us to reveal a temporal evolution like in steering tasks (e.g., Warren et al., 2001). These continuous dynamics suggest the visual system relies on time-varying optic flow (Burlingham and Heeger, 2020) to perceive a "future path" (Cutting et al., 1992; Li and Cheng, 2011), rather than extracting a momentary heading vector from instantaneous flow (Perrone and Stone, 1994; Stone and Perrone, 1997). Through sustained gaze stabilization, the brain moves beyond immediate transients, integrating foveal signals into a stable representation of the upcoming trajectory. Readers can experience this temporal build-up firsthand: maintaining a sustained lateral gaze while walking in an open space gradually induces a subtle, integrated drift that eventually becomes a clear departure from the intended path.

Active Steering vs. Heading Recovery

Navigation research distinguishes passive heading perception from active steering control (Goodridge et al., 2023; Powell et al., 2024; Wilkie and Wann, 2003). Traditional models assume explicit heading recovery (e.g., locating an instantaneous FoE) is required for steering, but our controller demonstrates that a trajectory can be maintained simply by nulling the error signal derived from retinal curl. Kim and Turvey (1999) and Wann and Swapp (2000) proposed a conceptually similar strategy that relies exclusively on sensitivity to retinal flow curvature in order to linearize the flow field. However, Saunders and Ma (2011) reported evidence against this pure linearization approach. Unlike this approach, our model does not depend on flow acceleration or curvature alone. Instead, it temporally integrates mean retinal curl, which is disambiguated by extra-retinal gaze variables. As derived in Appendix 1, raw curl becomes a precisely calibrated steering command only when scaled by gaze orientation, eye height, and translation speed (Frenz and Lappe, 2005). This gaze-scaled curl allows the visual system to continuously regulate the geometry of a stable "future path" without extracting a momentary heading vector (Tuhkanen et al., 2021; Wann and Swapp, 2000; Wilkie and Wann, 2006).

To validate this mechanism, we applied our controller to the paradigms of Wilkie and Wann (2003) (Experiment 3), successfully matching their reported empirical steering paths (Appendix 3). While they proposed a model integrating retinal flow with egocentric cues like visual direction (Rushton et al., 1998), our empirical findings demonstrate that flow dynamics can completely override egocentric signals. Even when an eccentric fixation point provided a strong, constant visual direction cue, cancelling the retinal curl eliminated the steering bias entirely, while our "over-cancelling" condition reversed its direction. This reversal contradicts a primary reliance on

the visual direction of the target. Instead, it suggests that in the absence of structural landmarks—and in the presence of reliable, large-field (>90°) optic flow—the visual system strongly favors gaze-scaled flow processing over visual direction (Wilkie and Wann, 2003).

Re-evaluating the Focus of Expansion (FoE)

The concept of the Focus of Expansion (FoE) was originally proposed by Gibson in the context of aircraft landings—a scenario characterized by minimal eye or head rotation (Gibson, 1950). However, the relevance of a translational head-centred FoE in natural contexts has been recently questioned (Matthis et al., 2022; Muller et al., 2023). While our results support the view that explicitly extracting a pure, de-rotated FoE is not required for steering, this does not render the FoE useless. The FoE likely remains highly relevant during high-speed locomotion, where gaze shifts are less frequent relative to the speed of travel, leading to a less disrupted flow field (Muller et al., 2023). Furthermore, even in complex scenes characterized by relative object motion, *pseudo-FoE* signals have been shown to drive heading biases (Layton and Fajen, 2016). We speculate that rather than subtracting global rotation to uncover a hidden FoE, the visual system exploits these uncorrected pseudo-FoE locations directly, combining this shifting retinal center of motion with extra-retinal signals and foveal curl to resolve the trajectory.

This perspective aligns closely with the neural model of MSTd proposed by Layton and Browning (2014), while our ring-attractor model extends this framework into a downstream parietal decoding mechanism. Layton and Browning (2014) suggested that MSTd integrates rotation directly rather than treating it as a nuisance component. In their model, MSTd hypercolumns represent self-motion simultaneously through the spirality (curl) of the most active units and the visuotopic location of this activity peak (the pseudo-FoE). Building upon this MSTd-like extraction of spiral motion (Duffy and Wurtz, 1991; Graziano et al., 1994), our neural model provides a biologically plausible bridge to motor control. We propose that downstream parietal regions, such as VIP (Schaafsma and Duysens, 1996) and 7a (Read and Siegel, 1997), to which MSTd projects (Born and Bradley, 2005), utilize a ring-attractor network to interpret this sensory output. In our model, standard center–surround (Mexican-hat) recurrent connectivity interacts with a weak straight-ahead prior and gaze-centered inhibition. These inherent push–pull dynamics effectively transform localized sensory inhibition into a global shift of the activity bump, naturally generating the psychophysical biases observed in our data. Furthermore, the network's recurrent dynamics act as a low-pass filter, smoothing high-frequency, gait-related temporal oscillations into stable trajectory estimates. Together, these frameworks demonstrate how standard cortical architectures can seamlessly transform uncorrected retinal flow into robust steering commands, without ever explicitly extracting a pure FoE.

Generalizability and Testable Predictions

Our results align with a growing consensus in sensorimotor research that incidental sensory signals often carry vital functional information (Rolfs and Schweitzer, 2022). Much like saccade-induced motion streaks facilitate gaze correction (Schweitzer and Rolfs, 2021), or oculomotor cycles actively format spatiotemporal visual input (Boi et al., 2017), retinal curl appears to be a systematic byproduct of gaze stabilization that the visual system exploits for navigation. Under this view, the biases observed in 3D environments are not failures of calculation, but rather the signature of a proportional controller shifting perceived heading toward a "null-curl" orientation.

This control-law perspective leads to several testable predictions. First, in environments with low visual texture or visibility (e.g., fog), the reliability of the curl signal should decrease, forcing the system to rely more heavily on "straight-ahead" internal priors and thus reducing the magnitude of the induced bias. Second, because our model relies on the precise calibration between extra-retinal signals and retinal curl, any perturbation of eye-position signals—whether through experimental manipulation or clinical conditions—should result in systematic steering errors that match the mis-scaling of the curl signal. Finally, while our focus was on ground-plane fixation, the

model predicts that any sustained gaze strategy that introduces structured rotation—such as tracking a moving agent or a point on a vertical wall—should generate predictable trajectory biases dictated by the specific geometry of the resulting foveal flow.

Conclusion

We conclude that the rotational component of optic flow (curl), generated during gaze stabilization, is an actively used signal to control heading. It acts as a navigational cue rather than noise, as evidenced by the elimination of steering biases when curl is experimentally cancelled. Our findings challenge the necessity of explicitly extracting the Focus of Expansion for online control of locomotion. The observed behaviors are supported by a neural model based on established properties of motion processing areas. The interaction between sensory flow inputs and internal priors within a recurrent network suffices to explain the gaze-contingent biases observed in our experimental data.

Methods

Participants

We tested 12 participants (five self-identified men and seven self-identified women), aged between 24 and 59 years (mean: 30, SD: 9), all with normal or corrected-to-normal vision. Except for one, all participants were naïve to the aims of the study and volunteered to take part. The study forms part of an ongoing research program approved by the Ethics Committee of the University of Barcelona (IRB 00003099) and conducted in accordance with the principles of the Declaration of Helsinki.

Displays and conditions

Participant motion was simulated as a translation parallel to the ground plane at a sustained walking speed of approximately 1 m/s. This motion incorporated characteristic bounce and swing components derived from a single gait profile (see Fig. S1 in the SI). This profile was recorded once by an independent individual wearing an HMD tracker while walking along the predefined experimental paths in a virtual environment, and then was applied to all participants to ensure stimulus consistency. While we acknowledge that using a standardized gait profile may introduce individual-level variability in perception, as the simulated motion may not align perfectly with each participant's unique gait dynamics, this approach was chosen to ensure that the reported perceptual biases are a robust consequence of the experimental variables (gaze direction and retinal curl) rather than artifacts of varying motion kinematics.

Each trial lasted between 11 and 12 s.

The ground plane consisted of a 50 × 50 m surface mapped with a naturalistic texture (see Fig. 1A) generated from simplex noise patterns whose spatial power spectrum followed a $1/f^2$ distribution. The temporal frequency of these patterns generated by simulated self-motion is consistent with the statistics of natural videos as described by (Dong and Atick, 1995) also following a 1/f-type temporal power spectra (exponent of 2). The textures were created in real time using OpenGL shaders within a custom Python program, designed for computational efficiency and allowing online manipulation of the texture in specific experimental conditions (see *Flow manipulation conditions* below).

The experiment was run on an Intel i7-based workstation (i7-9700F, Intel, Santa Clara, CA, USA) equipped with an NVIDIA GeForce RTX 2060 SUPER GPU. Images were rendered at 120 Hz with a resolution of 1920 × 1080 pixels and displayed monocularly via a PROPixx projector (VPIxx Technologies, Saint-Bruno, QC, Canada) onto a back-projection screen (2.03 m × 1.16 m), viewed from a distance of 1.0 m, resulting in a visual field of approximately 91º. The scale of the visual stimulus was precisely matched to the simulated environment by aligning the virtual camera's field of view with the physical geometry of the screen, ensuring a veridical 1:1 mapping between virtual and physical space.

To verify fixation, eye movements were recorded with a Pupil Labs Core (Berlin, Germany) eye tracker operating at 200 Hz. Trials were discarded if the median distance between gaze position and the fixation point exceeded 3º.

Path conditions

Participant trajectories could be either straight (length of about 6.5 m) or curved to the left or right (see Fig. 1B). The curved paths resulted in a final heading of ±45° relative to the initial heading (0°) in world coordinates. The three path types (straight, left, and right) were interleaved on a trial-by-trial basis.

Eccentricity conditions

In all trials, a fixation point (yellow dot in Figure 1A) was presented on the ground and remained fixed in world coordinates. Relative to the participant's initial position (x = 0, z = 0) and heading (0°), the fixation point could appear at one of five lateral positions or eccentricities, x = {-4, -2, 0, 2, 4} m, and was initially located 20 m ahead (see colored dots in Fig. 1B). As simulated self-motion progressed, the fixation point appeared to approach the observer, necessitating a gradual increase in gaze angle. Participants were not explicitly instructed to use specific eye or head movements to maintain fixation; instead, they were permitted to engage the head-eye system naturally to track the target. Under perfect fixation, the expected head/eye rotation rate increased for the largest eccentricity from about 0.4 º/s to 0.8°/s (see Fig. S2 in the SI). The eccentricity condition was randomized across trials.

Flow (curl) manipulation conditions

When fixating a stationary eccentric point while translating straight ahead, the retinal image contains an expected rotational component— referred to as curl—around the fovea. This effect is illustrated by the flow lines in Fig. 1A, which shows a snapshot of the retinal image consistent with walking straight ahead while fixating a point located to the left (positive curl). Figures 1C–E display the mean curl distributions for fixation points at varying eccentricities: leftward positions (−4 m, −2 m) produced positive curl, the center (0 m) resulted in near-zero curl, and rightward positions (2 m, 4 m) produced negative curl.

To compute the curl shown in Figs. 1C–E, we simulated experimental trials (videos downsampled to 800 × 452 px at 30 Hz) and computed optic flow using the Farnebäck algorithm (OpenCV). To ensure robust measurements and alleviate texture-dependent variability that can be introduced by the Farnebäck algorithm, we ran each trial 10 times using different textures always consistent with natural image statistics. For each frame t, we extracted the 2D image curl from the horizontal (f_x) and vertical (f_y) flow components:

$$\mathrm{curl}(x, y, t) = \frac{\partial f_y}{\partial x} - \frac{\partial f_x}{\partial y} \tag{1}$$

Beyond this algorithmic extraction, the retinal curl magnitude is geometrically determined by the observer's translational speed (v), eye height (h), and gaze orientation (pitch α and yaw ψ). As derived in Appendix 2, for a ground plane this relationship approximates to:

$$\mathrm{curl} \approx \frac{v \sin\psi \cos\alpha}{h} \tag{2}$$

We computed the mean image curl ω for each frame. Since the visual scene included gaitrelated oscillations (affecting the curl through both pitch and yaw), the raw computed curl was temporally smoothed (temporal window of 2.4 s). This processed signal was then used to make predictions based on the controller, ensuring robustness to high-frequency head movements. These values approximate the mean retinal curl expected under accurate fixation and define the *unaltered curl condition*.

In some conditions, this expected curl was counteracted (*cancelled curl condition*) by adding an equal and opposite rotational component centered on the fixation point. Figure 1C,E also show the corresponding curl histograms (lighter colors), which are centered at zero—matching the distribution obtained when fixating straight ahead in the direction of motion (mean zero curl, Figure 1D). In an additional set of trials, the imposed counter-rotation was doubled to produce an *overcancelled condition*, in which the curl was reversed beyond neutralization (histograms not shown in Figure 1).

Procedure

Participants continuously reported their perceived self-motion direction in the 3D scene while maintaining fixation on a target dot. Responses were collected via a custom rotative encoder (0.3° angular resolution) configured as an intuitive steering wheel interface, which participants reported was very easy to operate. To ensure high-frequency, non-blocking data acquisition, the device was interfaced through an Arduino Uno. We utilized a dedicated Python thread for direct serial communication, allowing for real-time polling of the encoder state without interfering with the primary stimulus rendering loop.

At the beginning of each session, participants completed a five-point calibration procedure for the eye tracker. Each session comprised 45 trials (3 trajectory types × 5 fixation positions × 3 flow manipulation conditions), and each participant completed a total of 10 sessions.

Computation of perceived path

To reconstruct the participant's estimated path from the angular response, we treated the reported heading angle θ_t (radians) as specifying the instantaneous lateral slope of the perceived trajectory at frame t. For each time step, we computed the incremental simulated forward displacement

$$\Delta z_t = z_t - z_{t-1}, \tag{3}$$

and converted the response angle into a lateral increment via

$$\Delta x_t = \tan\theta_t \Delta z_t. \tag{4}$$

The estimated lateral position was then obtained by forward integration starting from the initial position x_0:

$$x_t = x_0 + \sum_{k=1}^{t} \tan\theta_k \Delta z_k. \tag{5}$$

This yields the sample-wise reconstructed lateral trajectory x_t consistent with the participant's angular responses.

Control Model

We interpret the reported path as the output of a curl–driven feedback mechanism inspired by point attractor models used previously in steering control (Fajen and Warren, 2007; Wilkie and Wann, 2003). Let θ_t denote the heading direction (direction of forward velocity in world coordinates) and ψ_t the gaze direction. We consider their difference

$$\phi_t = \psi_t - \theta_t$$

as a instantaneous heading error. The mean optic–flow curl ω_t is taken as a sensory measurement of this discrepancy, so that $\phi_t \approx \omega_t / k_c$, with k_c being a curl–to–angle scale factor. This is motivated by the fact that curl increases whenever there is a discrepancy between θ and ψ. The observer controls heading by applying a yaw rate input $\omega_t = \dot{\theta}_t$:

$$\dot{\theta}_t = \omega_t = K_p \phi_t \approx K_p \frac{\omega_t}{k_c} \tag{6}$$

so that positive curl induces a leftward turn (increasing θ) and negative curl induces a rightward turn. If gaze is held fixed (e.g. steady fixation), $\dot{\psi}_t = 0$ and therefore

$$\phi_t = -\theta_t = -K_p \phi_t \tag{7}$$

This is a simple first-order linear differential equation, and the heading error decays exponentially with a correction time constant $1/K_p$:

$$\phi_t = \phi_0 e^{-K_p t} \tag{8}$$

Thus, heading is continuously steered toward gaze until curl vanishes. In our reconstruction we integrate this controlled heading forward in time to generate the perceived trajectory implied by the observed curl (see video S3 in the SI).

Trajectory fits

To test if perceived trajectories can be accounted for by the curl signal, we used the controller-defined in Equation 6 which we integrated forward with an explicit Euler step to obtain the estimated heading θ. With sampling interval Δt, constant forward speed (v), and initial state (θ_0, x_0, z_0) taken from the first sample of each trial, the discrete update is

$$\theta_{t+1} = \theta_t + \frac{K_p}{K_c} \overline{\omega}_t \Delta t$$
$$z_{t+1} = z_t + v \cos \theta_t \Delta t \quad x_{t+1} = x_t + v \sin \theta_t \Delta t \tag{9}$$

This generates a predicted 2D path $P_{\text{pred}}(K_p) = \{(z_t, x_t)\}_{t=1}^{T}$ from the observed mean curl (ω) sequence. We then estimate K_p and forward velocity (v) as free parameters by minimizing a time-warped path discrepancy (J) using Dynamic Time Warping (DTW)(Giorgino, 2009) on the (z, x) sequences between the predicted and perceived paths:

$$J(K_p) = \text{DTW}(P_{\text{obs}}, P_{\text{pred}}(K_p))$$

where

$$P_{\text{obs}} = \{(z_t^{\text{obs}}, x_t^{\text{obs}})\}_{t=1}^{T}$$

is the observed path.

We fix K_c (Equation 6) to 1, absorbing the unknown curl-to-angle scale into K_p.

We used two fitting approaches: (1) **independent** or **separate** fits, in which parameters were fitted independently for each combination of heading, gaze eccentricity, and retinal flow manipulation and (2) **join fits**, in which a single set of parameters is used across all heading and gaze eccentricity conditions but they were different for each flow manipulation. Each approach was tested with two parameters (K_p, and the forward velocity, v). The motivation for including forward velocity in the fit was to compensate for variations in the timing of perceived heading responses. For the separate fits, the number of parameters was 30 (for the 2-parameter approach, K_p and v). In contrast, the joined approach used only 2 parameters for all heading and eccentricity conditions.

In order to compare the different models, we employed information criteria adapted for distance-based model comparison. For each model, we computed a surrogate log-likelihood based on the normalized alignment cost:

$$\log \mathscr{L} = -\frac{1}{2}\left(\frac{D}{n}\right)$$

where D represents the total DTW distance and n the number of observations, with D/n representing the average alignment cost per observation. This normalization ensures appropriate scaling for information criteria computation. We then computed the Akaike Information Criterion as:

$$\mathrm{AIC} = 2k - 2\log \mathscr{L}$$

where k represents the number of parameters.

Data availability

All the data and code for the analysis are available at *https://osf.io/b37rg/*

Acknowledgements

We thanks Cristina de la Malla for her helpful comments and suggestions on the manuscript.

Additional information

Funding

JLM was supported by Grant PID2023-150081NB-I00 funded by MICIU/AEI/10.13039/501100011. KIZ was supported by fellowship PREP2023-001890 from MICIU.

Funding

Funder	Grant reference number	Author
MEC | Agencia Estatal de Investigación (AEI)	MICIU/AEI/10.13039/501100011	Konytessa I Zorpala

Author ORCID iDs

Joan López-Moliner: https://orcid.org/0000-0001-5040-8889

Additional files

Supplemental figures

Video S1. Retinal flow pattern during forward translation with fixation on a target (white dot) located to the left of the heading. For optimal viewing, the display should be centered relative to the observer with a field of view exceeding 50º (e.g., a 60 cm screen viewed from a distance of 60 cm).

Video S2. The same as Video S1, retinal flow pattern during forward translation and fixating a target (white dot) located to the left of the current path, but, the rotational component of the flow has been counteracted, so there is very little curl around the fovea. For optimal viewing, the display should be centered relative to the observer with a field of view exceeding 50º (e.g., a 60 cm screen viewed from a distance of 60 cm).

Video S3. Controller illustration: how heading is corrected to match gaze by applying equation 5 in the main text.

Appendix 1

Derivation of curl

Let's set up an eye-centered coordinate frame (X, Y, Z), where the eye is at the origin and the Z-axis is the line of sight (gaze direction) pointing directly at the fixation point.

- The observer moves with forward velocity v along the world heading.
- The fixation point is on the ground at 3D distance d from the eye and eye height is h.
- The gaze is offset from the heading by a yaw angle ψ horizontally and pitched downward by an angle α, where $\sin \alpha = \frac{h}{d}$.

- Projecting the forward velocity v into our eye-centered frame gives the eye's translational velocity (v_x, v_y, v_z):
- $v_x = -v \sin \psi$
- $v_y = -v \sin \alpha \alpha \cos \psi$
- $v_z = v \cos \alpha \cos \psi$

To find retinal flow, we need inverse depth $\frac{1}{Z}$ as a function of the image plane coordinates (x, y). We assume that the ground plane is perfectly horizontal in the world, but in our eye-centered frame (which is pitched down by α), the equation of the ground plane becomes: $Z \sin \alpha - Y \cos \alpha = h$

Dividing by Z and substituting the image coordinate $y = \frac{Y}{Z}$, we get the inverse depth profile of the ground:

$$\frac{1}{Z} = \frac{\sin \alpha - y \cos \alpha}{h} \tag{10}$$

Notice that looking vertically in the image plane changes the depth. The gradient of inverse depth at the fovea ($x = 0, y = 0$) is:

$$\frac{\partial}{\partial x}\left(\frac{1}{Z}\right) = 0 \tag{11}$$

$$\frac{\partial}{\partial y}\left(\frac{1}{Z}\right) = -\frac{\cos \alpha}{h} \tag{12}$$

Calculating the Curl of the translational flow

The standard optic flow equations for translation on the image plane (f_x, f_y) are:

$$f_x = \frac{-v_x + x v_z}{Z} \tag{13}$$

$$f_y = \frac{-v_y + y v_z}{Z} \tag{14}$$

The 2D curl at the fovea is defined as

$$C = \frac{\partial f_y}{\partial x} - \frac{\partial f_x}{\partial y} \tag{15}$$

Evaluating this at the origin ($x = 0, y = 0$):

$$\frac{\partial f_y}{\partial x} = -v_y \frac{\partial (1/Z)}{\partial x} = 0 \tag{16}$$

$$\frac{\partial f_x}{\partial y} = -v_x \frac{\partial (1/Z)}{\partial y} = -v_x \left(-\frac{\cos \alpha}{h}\right) = v_x \frac{\cos \alpha}{h} \tag{17}$$

Subtracting the two gives the translational curl:

$$C_{trans} = 0 - \left(v_x \frac{\cos \alpha}{h}\right) = -v_x \frac{\cos \alpha}{h} \tag{18}$$

Substituting our earlier $v_x = -v \sin \psi$ into equation Equation 18 we obtain the final expression for curl:

$$C_{\text{trans}} = v \cdot \sin \psi \frac{\cos \alpha}{h} \quad (18, 2)$$

Appendix 2

Neural Network Model of Gaze-Contingent Heading Bias

We present a neural model that provides a neurophysiologically plausible implementation of the controller. The model demonstrates that the bias emerges from the interaction between gaze-modulated visual flow processing and a weak straight-ahead prior in parietal heading representation, implemented through standard cortical connectivity. After presenting the model, we show different dynamic aspects of the model and reproduce the bias for the straight ahead trajectories looking at different eccentricities.

Local Motion Encoding

Visual motion is encoded by direction-selective units analogous to primate MT cortex, with 8 directional preferences:

$$\mathbf{E}_{\text{dirs}} = [\cos(\theta_k), \sin(\theta_k)]^T, \quad \theta_k = \frac{2\pi k}{8}, \quad k = 0, \ldots, 7 \quad (20)$$

The response of each direction channel at image-plane position $\mathbf{p}_i = (x_i, y_i)$ is given by rectified cosine tuning (Simoncelli and Heeger, 1998):

$$m_{i,k} = \max(0, \mathbf{v}_i \cdot \mathbf{E}_k) \quad (21)$$

where $\mathbf{v}_i$ is the local optical flow vector. In our implementation, optic flow $\mathbf{v}_i$ is computed, as before, using the Farnebäck algorithm as a substitute for early visual processing (V1/MT complex). This provides the motion signals that the visual system would extract through V1 to MT processing. The model itself begins with the input to the 8-sector MT-like directional encoding (Equation 21).

Curl Computation (MSTd)

To quantify the rotational (curl) component of optic flow around the current gaze position, we project local flow vectors onto the tangential direction of the circle centered at gaze. For each sampled location $\mathbf{p}_i$, we first compute its position relative to the gaze point $\mathbf{g}$, also encoded in image coordinates (x_i, y_i):

$$\mathbf{r}_i = \mathbf{p}_i - \mathbf{g} \quad (22)$$

This vector points from the gaze location to the sample point. The direction orthogonal to this vector corresponds to the local tangential direction of rotation around gaze. We obtain this unit tangential vector by rotating $\mathbf{r}_i$ by 90° and normalizing:

$$\hat{\mathbf{t}}_i = \frac{[-r_{i,y}, r_{i,x}]^T}{\|\mathbf{r}_i\|} \quad (23)$$

Finally, the contribution of the optic flow at that point to local rotational motion is given by the projection of the flow vector $\mathbf{v}_i$ onto this tangential direction:

$$\omega_i = \mathbf{v}_i \cdot \hat{\mathbf{t}}_i$$

Positive values indicate counterclockwise motion around gaze, and negative values indicate clockwise motion. The mean curl is computed as follows:

$$\bar{\omega} = \frac{1}{N_{\text{used}}} \sum_{i \in S} \omega_i$$

where S represents samples satisfying $\|\mathbf{r}_i\| > r_{\min}$ after trimming outliers. In our implementation, $N_{\text{total}} = 400$ samples are drawn within a gaze-centered region.

Ring Attractor Dynamics: Parietal Heading Representation

Heading direction θ is represented by a ring attractor mechanism consistent with previous studies (Zhang, 1996), with neural activity $x(\Phi, t)$ at preferred heading Φ evolving as:

$$\tau \frac{dx(\phi,t)}{dt} = -x(\phi,t) + \int W\left(\phi,\phi'\right) x\left(\phi',t\right) d\phi' + I_{\text{ext}}(\phi,t) \tag{24}$$

We implement these dynamics by using a discrete ring of $N = 181$ units with preferred headings Φ_j uniformly spaced between $[-\Phi_{\max}, \Phi_{\max}]$. The activity dynamics then follow:

$$x_j[t+1] = x_j[t] + \frac{\Delta t}{\tau}\left(-x_j[t] + \sum_{k=1}^{N} W_{jk} x_k[t] + I_{\text{ext},j}[t]\right) \tag{25}$$

The recurrent connectivity follows a standard Mexican hat profile (Ben-Yishai et al., 1995):

$$W_{jk} = A_e \exp\left(-\frac{d(\phi_j,\phi_k)^2}{2\sigma_e^2}\right) - A_i \exp\left(-\frac{d(\phi_j,\phi_k)^2}{2\sigma_i^2}\right) \tag{26}$$

where $d(\Phi_j, \Phi_k)$ is the circular distance between preferred headings. A_e and A_i were set to 1.9 and 1.0 respectively.

External Inputs

The external input $I_{\text{ext}}(\Phi, t)$ consists of two components:

$$I_{\text{ext},j}[t] = I_{\text{gaze},j}[t] + I_{\text{prior},j} \tag{27}$$

The gaze-modulated inhibitory input is:

$$I_{\text{gaze},j}[t] = -K_p \cdot \bar{\omega}(t) \cdot \exp\left(-\frac{d(\phi_j,\phi_g)^2}{2\sigma_\kappa^2}\right) \tag{28}$$

where Φ_g is the gaze direction in heading coordinates, K_p is the inhibitory gain, and σ_κ controls the width of the gaze-centered modulation, determining how broadly the inhibition spreads around the gaze direction Φ_g

The straight-ahead prior is:

$$I_{\text{prior},j} = I_0 \cdot \exp\left(-\frac{\phi_j^2}{2\sigma_p^2}\right) \tag{29}$$

I_0 is the prior strength and σ_p sets the width of the straight-ahead prior, with smaller values producing a sharper preference for $\Phi = 0$

Bias generation as Sensory–Prior Competition

Within this framework, the systematic bias opposite to gaze emerges from the interaction between the *gaze-centered inhibitory drive* and a *weak straight-ahead prior* within a stabilizing recurrent Mexican-hat network. A strong prior ($I_0 \gtrsim 0.25$) keeps the heading estimate near straight-ahead, while a weak prior $I_0 \approx 0.03$ allows sensory evidence to shift the attractor state. The symmetric recurrent dynamics maintain a coherent and stable activity bump. For leftward gaze ($\Phi_g < 0$) and forward motion ($\bar{\omega} > 0$): 1) Gaze-centered inhibition creates activity suppression at Φ_g; 2) With weak prior, activity bump shifts toward opposite side and 3) Recurrent dynamics translate inhibition into bump displacement.

The heading direction is decoded using population vector readout:

$$\hat{\theta} = \arg\left(\sum_{j=1}^{N} x_j e^{i\phi_j}\right) \tag{30}$$

where N is the number of ring units ($N = 181$). Unlike some heading perception models (Beintema et al., 2004; Lappe and Rauschecker, 1993), we do not apply sigmoid nonlinearities to neural activities prior to decoding but a linear readout. Gaze-contingent bias emerges even with linear decoding, suggesting it is a fundamental property of the network dynamics.

Parameter	Value range	Description
K_p	0.4–0.8	Inhibitory gain
I_0	0.03–0.24	Prior strength
σ_p	0.18–0.25	Prior width
σ_κ	0.02–0.12	Inhibition width
τ	60 ms	Membrane time constant
N	181	Ring population size
N_{samples}	400	Gaze-centered flow samples

Appendix 2 Table 1. Parameter values for reproducing experimental biases. σ values are expressed in fractions of azimuth range in pixels which was set to 360 in our simulations.

Simulation parameters

The specific parameter values used to produce Figure S16 are: $I_0 = 0.03$, $K_p = 0.4$, $\sigma_p = 0.18$, $\sigma_k = 0.12$.

Heading Estimation and 3D Path Reconstruction

The neural heading estimate θ_{px}, decoded from the ring attractor's peak activity, was converted from image coordinates to an ego-centric heading angle θ_{rad} using a pinhole camera model. Before converting to angular units, we smoothed the ring attractor's decoded signal to align with psychophysical evidence that complex motion is integrated over longer time scales than local

signals (Burr and Santoro, 2001). We used a temporal window of 2.4 seconds. Given a horizontal field of view (FOV_H = 100°) and an image width (W = 800 pixels), the camera's focal length in pixels was defined as:

$$f = \frac{W}{2\tan\left(\frac{FOV_H}{2}\right)}$$

The instantaneous heading in radians was then calculated by taking the inverse tangent of the pixel displacement relative to the principal point:

$$\theta_{\text{rad}} = \arctan\left(\frac{\theta_{px}}{f}\right)$$

To reconstruct the locomotor path in 3D space, we assumed a constant forward velocity v = 0.7 m/s. Under the assumption of a flat ground plane (fixed height y), the agent's position in world coordinates (X, Z) was updated via path integration. For each time step Δt = 0.033s, the position transition was defined as:

$$\begin{bmatrix} X_{t+1} \\ Z_{t+1} \end{bmatrix} = \begin{bmatrix} X_t \\ Z_t \end{bmatrix} + \begin{bmatrix} v \cdot \sin(\theta_{rad}) \\ v \cdot \cos(\theta_{rad}) \end{bmatrix} \Delta t$$

where Z represents the forward depth axis and X the lateral displacement. This way we were able to compare the reproduced path from the network model with the experimental data.

Relation to the controller (phenomenological model)

The neural implementation provides a mechanistic basis for the phenomenological relationship $\dot{\theta} = K_p \cdot \bar{\omega}(t)$ by demonstrating how rotational flow is converted into a change in perceived heading. In the neural model, K_p is not a single fixed parameter, but rather a dynamic property of the recurrent network. The shift in the activity bump (representing θ) is driven by the spatial asymmetry of the inputs. When the gaze-located inhibition suppresses the sensory evidence at the fixation point, it causes the stable bump of activity to "drift" toward the uninhibited regions of the sensory input. The speed of this drift is directly proportional to the strength of the inhibitory gain (K_p in Equation 28) and the magnitude of the local curl signal (ω). Thus, the network naturally performs the integration required by the controller: it transforms localized, gaze-contingent inhibition into a global heading shift that matches the $\dot{\theta} \propto \bar{\omega}$ relationship observed in participants.

Appendix 3

Controller fits to experimental data

We present the reported perceived headings alongside corresponding model fits derived from different approaches. These approaches vary based on parameter independence:

Separate Fits (Independent Parameters): Parameters are fitted independently for each combination of heading, gaze eccentricity, and retinal flow manipulation. Joint Fits (Common Parameters): A single set of parameters is used across all heading and gaze eccentricity conditions. Each approach was tested with two parameters (the controller gain, K_p, and the forward velocity, v). The initial motivation for including forward velocity in the fit was to compensate for variations in the timing of perceived heading responses. However, the resulting fitted velocity values (mean ± SD: v = 0.87 ± 0.12 m/s) were consistently close to the simulated physical speed during the steady phase (see Fig. 1 Figure Supplement 1).

Model Evaluation Metrics

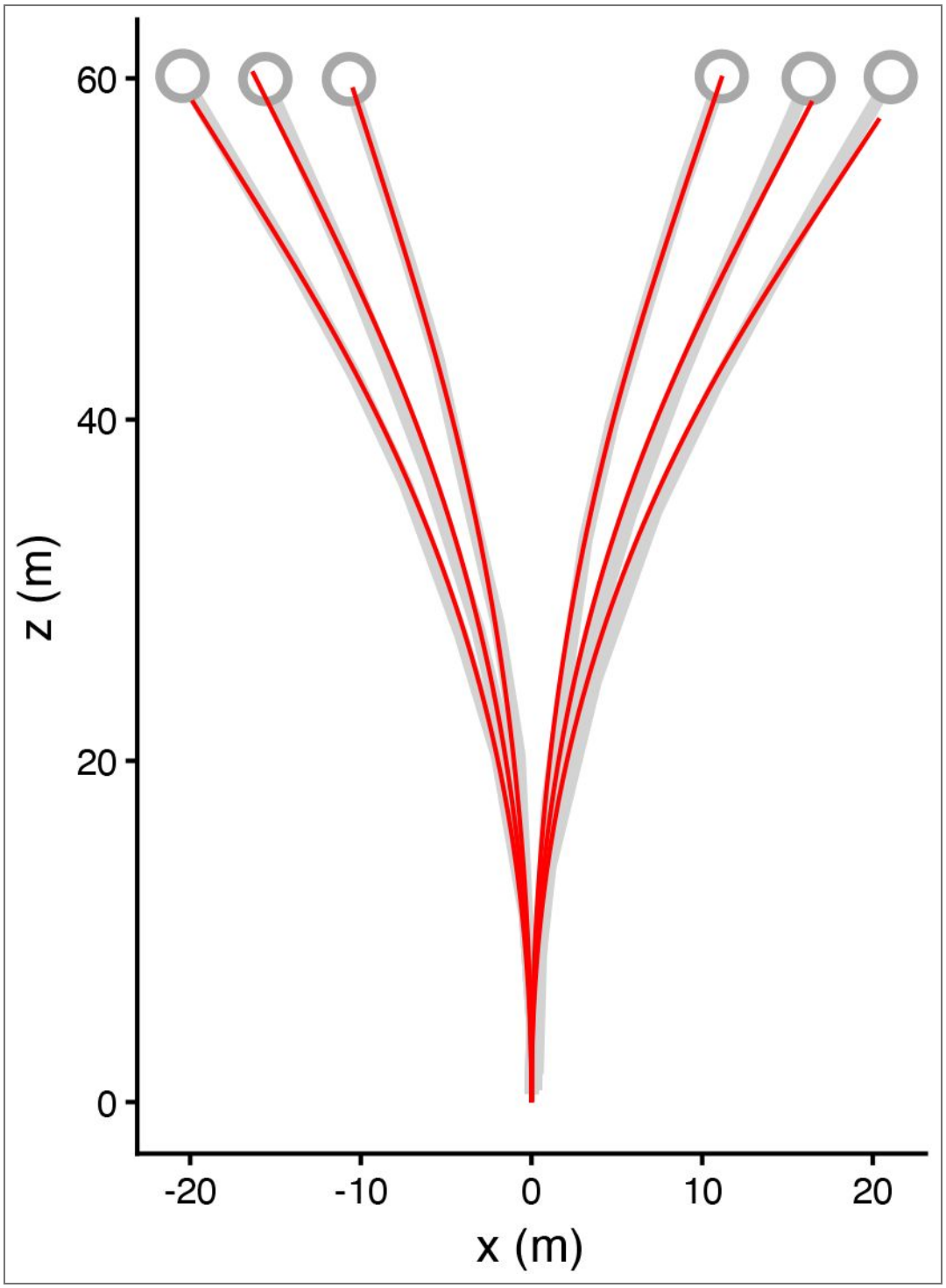
Model Type	Flow	Normalized Cost	Negative log-like	AIC
Separate (30 params)	unaltered	0.080	-3.54	67.08
Separate (28 params)	cancelled	0.066	-3.41	62.82
Separate (28 params)	overcancelled	0.077	-3.49	62.98
Join (2 param)	unaltered	0.379	-4.32	12.65
Join (2 param)	cancelled	0.252	-4.08	12.17
Join (2 param)	overcancelled	0.330	-4.22	12.44

Appendix 3 Table 1. Performance for the two fitting approaches (separate and join: models) in the different retinal flow conditions. Cost is the average 2D deviation per step between the fit and the observed heading.

- Figures from Figure 4 Fig. Suppl. 2 to Suppl. 7: separate fits
- Figures from Figure 4 Fig. Suppl. 8 to Suppl. 13: joint fits.

Controller fits to Exp 3 from Wilkie and Wann (2003)

To evaluate the generalizability of the curl-based controller, we fitted the controller to the human steering trajectories reported by Wilkie and Wann (2003). In order to obtain the initial image curl signals by simulating an observer kinematics over an 8-second duration on the same floor of our experiments with initial gaze eccentricities as in Wilkie and Wann (2003) and computed the mean curl using the generated videos of the simulations according to eq. 1. The controller transformed the instantaneous mean curl into a turn rate (ω) by adjusting the proportional gain (K_p) and the forward velocity (v) which minimized the 2D distance between the simulated paths and the steering trajectories by Wilkie and Wann (2003). As shown in S17 in the SI, the model accurately captures the steering behavior across different target eccentricities, with the predicted paths (red) closely aligning with the steering trajectories (grey). Table S3 in the SI shows the fitted parameters and the final lateral position which always ended within the target. The controller gain K_p was very similar in all paths indicating that the differences were caused by the curl signal. The fitted speed was very close to the experimental speed (8 m/s) used in Wilkie and Wann (2003).



Appendix 3 Fig 1. Reproduction of steering paths towards targets. Response of the controller model (red lines) to the conditions introduced in Experiment 3 of Wilkie and Wann 2003.Targets were placed 60 m ahead at 10, 14, and 18 to the left and right (grey circles). The grey lines denote an approximation of the observer paths reported in Wilkie and Wann (2003). The simulated speed in their study was 8 m/s

Target initial eccentricity (º)	Kp	Vel (m/s)	Final x (m)
-18º	3.54	8.00	-19.9
-14º	3.57	8.03	-16.4
-10º	3.21	7.7	-10.5
10º	3.38	7.8	11.2
14º	3.69	7.82	16.4
18º	3.68	7.90	20.4

Appendix 3 Table 2. Parameters of the controller to successfully steer to the 6 targets shown in figure S17. The controller speed was very close to the simulated speed (8 m/s) in (Wilkie and Wann, 2003). The final lateral position always ended within the target dimensions (2 m width).

References

Angelaki DE, Hess BJ (2005) Self-motion-induced eye movements: Effects on visual acuity and navigation. *Nature Reviews Neuroscience* **6**:966-976 https://doi.org/10.1038/nrn1804 | PubMed

Banks MS, Ehrlich SM, Backus BT, Crowell JA (1996) Estimating heading during real and simulated eye movements. *Vision research* **36**:431-443 https://doi.org/10.1016/0042-6989(95)00122-0 | PubMed

Bardy B, Warren W, Kay B (1999) The role of central and peripheral vision in postural control during walking. *Perception & Psychophysics* **61**:1356-1368 https://doi.org/10.3758/BF03206186 | PubMed

Beintema JA, van den Berg AV (1998) Heading detection using motion templates and eye velocity gain fields. *Vision research* **38**:2155-2179 https://doi.org/10.1016/S0042-6989(97)00428-8 | PubMed

Beintema JA, van den Berg AV, Lappe M (2004) Circular Receptive Field Structures for Flow Analysis and Heading Detection. In: Vaina LM, Beardsley SA, Rushton SK (Eds). *Optic Flow and Beyond* Springer Netherlands. pp. 223-248 https://doi.org/10.1007/978-1-4020-2092-6_10

Ben-Yishai R, Bar-Or RL, Sompolinsky H (1995) Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences* **92**:3844-3848 https://doi.org/10.1073/pnas.92.9.3844 | PubMed

Boi M, Poletti M, Victor JD, Rucci M (2017) Consequences of the oculomotor cycle for the dynamics of perception. *Current Biology* **27**:1268-1277 https://doi.org/10.1016/j.cub.2017.03.034 | PubMed

Born RT, Bradley DC (2005) Structure And Function Of Visual Area Mt. *Annual Review of Neuroscience* **28**:157-189 https://doi.org/10.1146/annurev.neuro.26.041002.131052 | PubMed

Bremmer F, Churan J, Lappe M (2017) Heading representations in primates are compressed by saccades. *Nature Communications* **8**:920 https://doi.org/10.1038/s41467-017-01021-5 | PubMed

Britten KH (2008) Mechanisms of Self-Motion Perception. *Annual Review of Neuroscience* **31**:389-410 https://doi.org/10.1146/annurev.neuro.29.051605.112953 | PubMed

Burlingham CS, Heeger DJ (2020) Heading perception depends on time-varying evolution of optic flow. *Proceedings of the National Academy of Sciences* **117**:33161-33169 https://doi.org/10.1073/pnas.2022984117 | PubMed

Burr DC, Santoro L (2001) Temporal integration of optic flow, measured by contrast and coherence thresholds. *Vision Research* **41**:1891-1899 https://doi.org/10.1016/S0042-6989(01)00072-4 | PubMed

Calow D, Lappe M (2008) Efficient encoding of natural optic flow. *Network: Computation in Neural Systems* **19**:183-212 https://doi.org/10.1080/09548980802368764 | PubMed

Cutting JE (1986) *Perception with an eye for motion* Cambridge: MIT Press.

Cutting JE, Springer K, Braren PA, Johnson SH (1992) Wayfinding on foot from information in retinal, not optical, flow. *Journal of Experimental Psychology: General* **121**:41 https://doi.org/10.1037/0096-3445.121.1.41 | PubMed

Dong DW, Atick JJ (1995) Statistics of natural time-varying images. *Network: Computation in Neural Systems* **6**:345-358 https://doi.org/10.1088/0954-898X_6_3_003

Duffy CJ, Wurtz RH (1991) Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology* **65**:1329-1345 https://doi.org/10.1152/jn.1991.65.6.1329 | PubMed

Fajen BR, Warren WH (2007) Behavioral dynamics of intercepting a moving target. *Experimental brain research* **180**:303-19 https://doi.org/10.1007/s00221-007-0859-6 | PubMed

Freeman TCA, Champion RA, Warren PA (2010) A bayesian model of perceiving head-centered velocity during smooth pursuit eye movement. *Current Biology* **20**:757-762 https://doi.org/10.1016/j.cub.2010.02.059 | PubMed

Frenz H, Lappe M (2005) Absolute travel distance from optic flow. *Vision research* **45**:1679-1692 https://doi.org/10.1016/j.visres.2004.12.019 | PubMed

Gibson JJ (1950) *The perception of the visual world* Houghton Mifflin.

Giorgino T (2009) Computing and visualizing dynamic time warping alignments in r: The dtw package. *Journal of statistical Software* **31**:1-24 https://doi.org/10.18637/jss.v031.i07

Glennerster A, Hansard ME, Fitzgibbon AW (2001) Fixation could simplify, not complicate, the interpretation of retinal flow. *Vision research* **41**:815-834 https://doi.org/10.1016/s0042-6989(00)00300-x | PubMed

Goodridge CM, Billington J, Markkula G, Wilkie RM (2023) Error accumulation when steering toward curves. *Journal of experimental psychology: human perception and performance* **49**:821 https://doi.org/10.1037/xhp0001101 | PubMed

Graziano MS, Andersen RA, Snowden RJ (1994) Tuning of MST neurons to spiral motions. *Journal of Neuroscience* **14**:54-67 https://doi.org/10.1523/JNEUROSCI.14-01-00054.1994 | PubMed

Grigo A, Lappe M (1999) Dynamical use of different sources of information in heading judgments from retinal flow. *Journal of the Optical Society of America A* **16**:2079-2091 https://doi.org/10.1364/JOSAA.16.002079 | PubMed

Heeger DJ, Jepson AD (1992) Subspace methods for recovering rigid motion I: Algorithm and implementation. *International Journal of Computer Vision* **7**:95-117 https://doi.org/10.1007/BF00128130

Kaminiarz A, Schlack A, Hoffmann K-P, Lappe M, Bremmer F (2014) Visual selectivity for heading in the macaque ventral intraparietal area. *Journal of Neurophysiology* **112**:2470-2480 https://doi.org/10.1152/jn.00410.2014 | PubMed

Kim N-G, Turvey MT (1999) Eye Movements and a Rule for Perceiving Direction of Heading. *Ecological Psychology* **11**:233-248 https://doi.org/10.1207/s15326969eco1103_3

Koenderink JJ, van Doorn AJ (1981) Exterospecific component of the motion parallax field. *Journal of the Optical Society of America* **71**:953-957 https://doi.org/10.1364/JOSA.71.000953 | PubMed

Lappe M (1998) A model of the combination of optic flow and extraretinal eye movement signals in primate extrastriate visual cortex: Neural model of self-motion from optic flow and extraretinal cues. *Neural Networks* **11**:397-414 https://doi.org/10.1016/S0893-6080(98)00013-6 | PubMed

Lappe M, Rauschecker JP (1993) A neural network for the processing of optic flow from ego-motion in man and higher mammals. *Neural Computation* **5**:374-391 https://doi.org/10.1162/neco.1993.5.3.374

Layton OW, Browning NA (2014) A unified model of heading and path perception in primate MSTd. *PLoS computational biology* **10**:e1003476 https://doi.org/10.1371/journal.pcbi.1003476 | PubMed

Layton OW, Fajen BR (2016) Sources of bias in the perception of heading in the presence of moving objects: Object-based and border-based discrepancies. *Journal of Vision* **16**:9-9 https://doi.org/10.1167/16.1.9 | PubMed

Li L (2025) Visual Perception of Self-Motion. *Annual Review of Vision Science* **11**:447-474 https://doi.org/10.1146/annurev-vision-121423-013200 | PubMed

Li L, Cheng JC (2011) Perceiving path from optic flow. *Journal of Vision* **11**:22-22 https://doi.org/10.1167/11.1.22 | PubMed

Li L, Warren WH (2002) Retinal flow is sufficient for steering during observer rotation. *Psychological Science* **13**:485-491 https://doi.org/10.1111/1467-9280.00486 | PubMed

Li L, Warren WHJ (2000) Perception of heading during rotation: Sufficiency of dense motion parallax and reference objects. *Vision research* **40**:3873-94 https://doi.org/10.1016/s0042-6989(00)00196-6 | PubMed

Longuet-Higgins HC, Prazdny K (1980) The interpretation of a moving retinal image. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **208**:385-397 https://doi.org/10.1098/rspb.1980.0057 | PubMed

Matthis JS, Muller KS, Bonnen KL, Hayhoe MM (2022) Retinal optic flow during natural locomotion. *PLOS Computational Biology* **18**:e1009575 https://doi.org/10.1371/journal.pcbi.1009575 | PubMed

Matthis JS, Yates JL, Hayhoe MM (2018) Gaze and the control of foot placement when walking in natural terrain. *Current Biology* **28**:1224-1233 https://doi.org/10.1016/j.cub.2018.03.008 | PubMed

Muller KS, Matthis J, Bonnen K, Cormack LK, Huk AC, Hayhoe M (2023) Retinal motion statistics during natural locomotion. *eLife* **12**:e82410 https://doi.org/10.7554/eLife.82410 | PubMed

Perrone JA, Stone LS (1994) A model of self-motion estimation within primate extrastriate visual cortex. *Vision research* **34**:2917-2938 https://doi.org/10.1016/0042-6989(94)90060-4 | PubMed

Powell NV, Marshall X, Diaz GJ, Fajen BR (2024) Coordination of gaze and action during high-speed steering and obstacle avoidance. *PLoS one* **19**:e0289855 https://doi.org/10.1371/journal.pone.0289855 | PubMed

Powell NV, Panfili D, Oh Y, Matthis JS, Hayhoe M (2026) The role of retinal flow in walking. *bioRxiv* https://doi.org/10.64898/2026.02.05.704053

Read HL, Siegel RM (1997) Modulation of responses to optic flow in area 7a by retinotopic and oculomotor cues in monkey. *Cerebral cortex (New York, NY: 1991)* **7**:647-661 https://doi.org/10.1093/cercor/7.7.647 | PubMed

Rieger JH, Toet L (1985) Human visual navigation in the presence of 3-D rotations. *Biological Cybernetics* **52**:377-381 https://doi.org/10.1007/BF00449594 | PubMed

Rolfs M, Schweitzer R (2022) Coupling perception to action through incidental sensory consequences of motor behaviour. *Nature Reviews Psychology* **1**:112-123 https://doi.org/10.1038/s44159-021-00015-x

Royden CS (1994) Analysis of misperceived observer motion during simulated eye rotations. *Vision research* **34**:3215-3222 https://doi.org/10.1016/0042-6989(94)90085-X | PubMed

Royden CS, Banks MS, Crowell JA (1992) The perception of heading during eye movements. *Nature* **360**:583-585 https://doi.org/10.1038/360583a0 | PubMed

Royden CS, Crowell JA, Banks MS (1994) Estimating heading during eye movements. *Vision research* **34**:3197-3214 https://doi.org/10.1016/0042-6989(94)90084-1 | PubMed

Rushton SK, Harris JM, Lloyd MR, Wann JP (1998) Guidance of locomotion on foot uses perceived target location rather than optic flow. *Current Biology* **8**:1191-1194 https://doi.org/10.1016/S0960-9822(07)00492-7 | PubMed

Saunders JA, Ma K-Y (2011) Can observers judge future circular path relative to a targetfrom retinal flow?. *Journal of Vision* **11**:16-16 https://doi.org/10.1167/11.7.16

Schaafsma SJ, Duysens J (1996) Neurons in the ventral intraparietal area of awake macaque monkey closely resemble neurons in the dorsal part of the medial superior temporal area in their responses to optic flow patterns. *Journal of Neurophysiology* **76**:4056-4068 https://doi.org/10.1152/jn.1996.76.6.4056 | PubMed

Schweitzer R, Rolfs M (2021) Intrasaccadic motion streaks jump-start gaze correction. *Science Advances* **7**:eabf2218 https://doi.org/10.1126/sciadv.abf2218 | PubMed

Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. *Vision research* **38**:743-761 https://doi.org/10.1016/S0042-6989(97)00183-1 | PubMed

Stone LS, Perrone JA (1997) Human heading estimation during visually simulated curvilinear motion. *Vision research* **37**:573-590 https://doi.org/10.1016/S0042-6989(96)00204-0

Tuhkanen S, Pekkanen J, Wilkie RM, Lappi O (2021) Visual anticipation of the future path: Predictive gaze and steering. *Journal of Vision* **21**:25-25 https://doi.org/10.1167/jov.21.8.25 | PubMed

van den Berg AV (1993) Perception of heading. *Nature* **365**:497-498 https://doi.org/10.1038/365497a0 | PubMed

van den Berg AV (1992) Robustness of perception of heading from optic flow. *Vision research* **32**:1285-1296 https://doi.org/10.1016/0042-6989(92)90223-6 | PubMed

van den Berg AV, Beintema JA (2000) The mechanism of interaction between visual flow and eye velocity signals for heading perception. *Neuron* **26**:747-752 https://doi.org/10.1016/S0896-6273(00)81210-6 | PubMed

van den Berg AV, Brenner E (1994) Why two eyes are better than one for judgements of heading. *Nature* **371**:700-702 https://doi.org/10.1038/371700a0 | PubMed

Wann J, Land M (2000) Steering with or without the flow: Is the retrieval of heading necessary?. *Trends in cognitive sciences* **4**:319-324 https://doi.org/10.1016/S1364-6613(00)01513-8 | PubMed

Wann JP, Swapp DK (2000) Why you should look where you are going. *Nature neuroscience* **3**:647-648 https://doi.org/10.1038/76602 | PubMed

Warren W (1998) The state of flow. In: Watanabe T (Ed). *High-level motion processing: Computational, neurobiological, and psychophysical perspectives* MIT Press. pp. 219-230

Warren WH, Hannon DJ (1990) Eye movements and optical flow. *Josa a* **7**:160-169 https://doi.org/10.1364/JOSAA.7.000160 | PubMed

Warren WH, Hannon DJ (1988) Direction of self-motion is perceived from optical flow. *Nature* **336**:162-163 https://doi.org/10.1038/336162a0

Warren WHJ, Kay BA, Zosh WD, Duchon AP, Sahuc S (2001) Optic flow is used to control human walking. *Nature neuroscience* **4**:213-6 https://doi.org/10.1038/84054 | PubMed

Warren WH, Morris MW, Kalish M (1988) Perception of translational heading from optical flow. *Journal of Experimental Psychology: Human Perception and Performance* **14**:646 https://doi.org/10.1037/0096-1523.14.4.646 | PubMed

Wilkie RM, Wann JP (2006) Judgments of path, not heading, guide locomotion. *Journal of Experimental Psychology: Human Perception and Performance* **32**:88-96 https://doi.org/10.1037/0096-1523.32.1.88 | PubMed

Wilkie RM, Wann JP (2003) Controlling steering and judging heading: Retinal flow, visual direction, and extraretinal information. *Journal of Experimental Psychology: Human Perception & Performance* **29**:363-378 https://doi.org/10.1037/0096-1523.29.2.363 | PubMed

Xie M, Li L (2025) Distinct temporal dynamics of judging scene-relative object motion and estimating heading from optic flow. *Journal of Vision* **25**:10-10 https://doi.org/10.1167/jov.25.13.10 | PubMed

Zhang K (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience* **16**:2112-2126 https://doi.org/10.1523/JNEUROSCI.16-06-02112.1996 | PubMed

Peer reviews

Reviewer #1 (Public review):

Summary:

This carefully executed study uncovers the functional relevance of curl signals that impinge on the retina every time an observer's gaze direction and movement direction are not aligned. This finding is important, highlighting the functional role of an abundant incidental signal (curl in retinal motion) that has thus far believed to be a nuisance that needs to be filtered out of the retinal motion stream. As such, the study forms an important contribution to the emerging recognition that incidental sensory signals are not a challenge to the sensorimotor system, but contain functionally relevant and effectively used visual signals. The study's evidence is compelling: A combination of psychophysical experiments and critical manipulations, control theory and neural modeling makes an internally consistent and biologically plausible case for the role of curl signals in estimating heading direction. The experimental and modeling results clearly go beyond previous studies and significantly advance our understanding of vision-based navigation.

Strengths:

The study has its strengths in the combination of psychophysical experiments and critical manipulations, control theory and neural modeling, which together make an internally consistent and biologically plausible case for the role of curl signals in estimating heading direction.

This study uncovers the functional relevance of curl signals that occur on the retina when an observer is moving and gaze is not straight ahead. The experimental and modeling results clearly go beyond previous studies and significantly advance our understanding of vision-based navigation.

Another clear strength is that the study uses tightly controlled experimental manipulation to provide strong test cases for the hypothesis that curl is used for visual navigation. These conditions are important to constrain the proposed model (and future models) of heading control.

The modeling is very clearly described and the modeling and analysis code is published and freely available. The authors go beyond a back-of-the-envelope control model and show how it might be implemented at the neural-circuit level. The model is biologically plausible.

Weaknesses:

I see no major weaknesses of the study. I expect it to inspire future research that extends these findings to a wider range of visual environments (including walking in natural scenes), motion speeds and kinds of movements.

Comments on revised version.

I have no additional comments for the authors.

https://doi.org/10.7554/eLife.110770.2.sa3

Reviewer #2 (Public review):

This study examines how curl in the retinal flow field can be used as a control variable for estimating and controlling the heading of a moving observer. The basic idea (which is not entirely new, see Matthis et al. 2022) is that translation along a path with eccentric gaze

(meaning that the subject is not heading toward the point they are looking at) produces a pattern of optic flow on the retina with a rotational component around the point of fixation (which can be captured by the mathematical "curl" operator). The sign and magnitude of retinal curl varies with heading relative to the point of fixation, such that curl can be used as a control variable to steer rightward or leftward to move toward the fixated target. The authors perform behavioral experiments and show that there are biases in perceived heading that seem to be largely governed by retinal curl. They also show that a simple controller model can use curl to steer toward a target, and they provide a neural network model that provides a biologically-plausible implementation of the controller (although there are some questions about that).

There is a core of interesting work here that I think can be important to the field. However, there is a lack of clarity on several important fronts, including design of the behavioral experiments, presentation of the behavioral data, conceptual framing of what curl can and cannot do, etc. Equally importantly, the manuscript is not written in a manner that will make it accessible to most vision scientists. I consider myself to be pretty knowledgeable about optic flow, and I had to read most of the manuscript 3 or 4 times to be able to understand the bulk of it. And my experience is that most vision scientists do not understand optic flow well, so I fear that most of the readers that the authors should want to reach would struggle to understand the work. As written, this is mainly going to make an impact on a handful of optic flow gurus. Thus, this manuscript is going to need a major overhaul to clarify important issues and make this more accessible.

Major issues:

(1) The manuscript contains inconsistent, if not misleading, messaging about what information retinal curl does, and does not, provide regarding heading estimation. In the Abstract, the authors state: "We propose an alternative: the visual system utilizes retinal curl directly to estimate heading, rendering the explicit recovery of the FOE unnecessary." Based on my understanding of the rest of the manuscript, I find this statement to be a misrepresentation for two main reasons:

a. To "directly estimate heading" relative to what? When not qualified, most people interpret "heading" to mean an observer's heading relative to the world (or some allocentric reference frame). But retinal curl only gives information about an observer's heading relative to the point on which their eyes are fixated. Moreover, that point of fixation will change every few hundred milliseconds in natural viewing, so the retinal curl will change with each new fixation even as heading relative to the world remains unchanged. So, I think most readers would grossly misinterpret the claim that retinal curl can be used "directly to estimate heading". Indeed, in the authors' controller model, the initial heading needs to be given and then the controller can work. But from where does the visual system get the initial heading, since it does not come from curl? These issues are left hanging. Thus, while curl can provide a very useful input for steering toward a fixated target, other signals are needed to estimate heading relative to the world. This has to be made much clearer early on, and a conceptual schematic diagram might help. Also, the authors generally do not specify the reference frame of the variables they are talking about, leaving lots of room for misinterpretations. It should be clear each time they are talking about a variable, such as heading, whether it is relative to the fixation target, body, world, etc.

b. It seems to me that retinal curl will depend on other variables, in addition to heading relative to the fixation target. For example, it seems to me that the magnitude of retinal curl will depend on self-motion speed, the depth structure of the scene, the angle of elevation of the fixated target, and perhaps others. This is not discussed at all, and many readers would get the misguided impression that there is a 1:1 mapping from curl to heading (relative to fixation). If I am right that this is not correct, it means that retinal curl can tell the observer whether to steer right or left to move toward the fixated target, but it cannot tell them how

much to steer. Indeed, in the authors' controller model, there is a free parameter that calibrates curl to angle. It makes sense that this works to fit trajectory data that are given from a fixed environment, but it is unclear how the brain would use retinal curl to control steering when these other variables are uncertain or changing unpredictably. Moreover, how does the system change the mapping from curl to steering command as the location of fixation changes relative to the current heading? These are issues that need to be brought up in framing the problem and discussed at some length. If the authors can show mathematically that retinal curl is only dependent on heading (relative to fixation) and not any of these other variables, it would be very valuable to show the equations for this relationship.

(2) The description of the behavioral experiment and presentation of behavioral data leaves a lot to be desired.

a. First, it is stated (line 158) that "Participants continuously reported their perceived direction of self-motion while maintaining fixation on the yellow dot." Again, reference frame is completely unspecified. Participants were reporting their perceived heading relative to what? The fixation target? The world? What exactly were the instructions given to the subjects to perform the task? Based on the description of how perceived paths are computed (line 166-), it seems to be presumed that subjects are reporting their heading relative to the world because those angles are then converted into x and z coordinates in what I presume is a world-centered reference frame. But how do we know that subjects are accurately reporting their heading relative to the world? What if they are biased in their reports by the location of the fixation target relative to the scene, or by some other reference signal? Is it possible for the authors to rule out the possibility that perceptual biases seen in the unaltered curl condition result from observers not fully adopting the assumed reference frame of the task? If this cannot be firmly excluded, it seems to create problems for the rest of the study.

b. I also feel that there is a mismatch between what the behavioral task requires and what the controller model does. Subjects are apparently asked to report their heading relative to the world, but the controller model only controls their heading relative to the point that they are fixating. I understand how this is resolved in the model, but I think this type of distinction is buried and will not be apparent to most readers. Again, the reference frames of what is being measured and controlled need to be specified explicitly in all parts of the paper, and the authors needs to explain how the system would combine curl-based control with some other measures of (at least initial) heading for world-centered heading to be computed. All of the assumptions need to be clearly specified.

c. Second, I found it frustrating that the authors never present raw perceptual data from the observers. Rather, in Figure 2, we see reconstructed trajectories that are perfectly smooth with no indications of noise whatsoever. Since these paths are computed from the perceptual reports, there must be some noise inherent in them. The figures should represent this uncertainty somehow, and it should be explained how these perfectly smooth trajectories are obtained.

(3) "...the magnitude of retinal curl in the fovea can specify the body trajectory relative to gaze (Matthis et al., 2022)." The main idea put forward by the authors here seems to overlap heavily with this statement that they attribute to Matthis et al. 2022. While I think this paper still adds importantly to the topic, the authors do not discuss how their findings are different from those of Matthis et al. 2022, why they are an important extension, etc. Readers should not have to go read this other paper to have any idea how the present findings are placed in importance relative to the literature.

(4) The analysis and treatment of eye movements is extremely weak. The authors discarded trials for which gaze deviated from the fixation point by more than 3 degrees (which is a LOT given that the eye speeds are generally in the neighborhood of 0.5 deg/sec), and they provide

basic stats on the distribution of positions. But this largely misses the point: it is not small position errors that are likely to matter, but rather velocity errors. Even a small amount of retinal slip of the target while it is being pursued will cause image motion that is going to alter the optic flow field around the fixation target. So, for example, the retinal curl field may no longer be centered on the fixation target. How do we know that some of the perceptual biases are not influenced by image motion resulting from imperfect tracking of the fixation target? This needs to be analyzed and discussed.

(5) I found the sections of text comparing the separate and joined fits (starting line 287) to be a bit too rosy. The authors show the separate fits in the main text, and it is not very surprising that these fits are good given that the model has 30 parameters, and these data are pretty low dimensional. The authors only show the joined fits in the supplement, and they say that they are almost as good as the separate fits (indeed they are better in a model comparison sense, but this is 30 parameters vs. 2 parameters). However, when I look at the fits of the joined model in the supplement, I don't find them to be very impressive. In particular, the model grossly misses the data for the straight paths for several subjects (e.g., id5, id6, id8, id10). And fitting the straight paths would presumably be easiest. This implies that the joined model is really missing something and that fitting the curved paths interacts strongly with fitting the data for different fixation target locations on the straight path. I think that the authors should discuss the results a bit more soberly and tone down their conclusions here.

(6) The section of the paper on neural simulations (starting line 387) has a few weaknesses. First, why are only straight paths simulated here? This does not seem to provide a very rigorous test of the model. Second, it is awkward that the simulation results are presented in units of pixels, rather than degrees. Third, the authors seem to downplay the fact that the neural estimates of heading seem to oscillate rather wildly (over a range of hundreds of pixels, whatever that means, see especially Fig. S16). It was far from clear to me how an estimate of heading with these large oscillations is useful. It would seem to require that heading estimates are integrated over substantial lengths of time to be reliable. It was therefore unclear how the model produces such smooth paths from these oscillating estimates.

Comments on revised version.

Overall, the authors have done a responsible job of responding to the comments of my previous review, and the manuscript is substantially improved. There are a few points on which I still do not completely agree with the authors, and I think these are important to document for the record:

(1) Introduction: "Pure visual decomposition should function regardless of 3D depth or whether the rotation stems from an active eccentric fixation." Perhaps in a world of noiseless perfect computation, this might be true. But I generally disagree. When there is more depth structure in an environment, then translation of the observer is generally going to create greater motion parallax. That is a fact that I don't think can be disputed. And greater motion parallax should help to decompose optic flow into components related to translation and rotation (the latter of which is not depth dependent), especially when there is noise in estimating location motion vectors.

(2) Related to point #9 of my previous review: I had asked why the authors believed that retinal curl was computed in area MSTd. Their response is that previous studies (i.e., Graziano et al. 1994) show selectivity to spiral motion stimuli in MSTd. That is true, but those studies typically placed the spiral stimulus centered on the MSTd receptive field, hence they were not presenting something like retinal curl as defined here. So, I think it is still an open question as to where in the brain retinal curl is encoded, and from which areas it would be possible to decode retinal curl from population responses.

(3) Related to point #10 of my previous review: I had asked about biological plausibility of the gaze-centered inhibition signal in the model. The authors' response is that parietal neurons show gain fields in which response depends (usually monotonically) on eye position. This is true, but it is not a trivial jump from gain fields in individual neural responses to a gaze-centered inhibition signal, and I think the authors should have been more forthcoming about the lack of an established neural signal that directly signals what they want in their model.

(4) The authors point out that the perceptual biases they measure take a few seconds to emerge and they attribute this to temporal integration. But in their curl manipulations, they temporally average over a 2.4 second window in computing the curl signals that they use to cancel or over-cancel curl. So, it is not clear whether some of the delay in the behavioral effects might result from their computations.

(5) Related to point #13 of my previous review: I had asked about empirical evidence for the assumption of a relationship between the heading preferences of MSTd neurons and their receptive field locations. In response, the authors state that such a relationship is built into the Layton and Browning (2014) model. While that is a precedent, citing another model as a response to a question about empirical evidence is not a convincing response. If there is no empirical evidence to support such a relationship, it would have been better for the authors to acknowledge this.

Given the way that the eLife review model works, it is not necessary for the authors to address these comments, but I think they should be included in the public review record.

https://doi.org/10.7554/eLife.110770.2.sa2

Reviewer #3 (Public review):

Major strengths include the use of realistic retinal motion recorded during virtual walking, an elegant manipulation of curl, converging behavioral and modeling evidence, and grounding in control theory. This provides a novel and important contribution to our understanding of how the brain processes motion information and intuition about how that information might be used to guide steering. In addition, they provide a computational mechanism by which retinal flow curl can be used as a control signal.

The revised ms has been strengthened by more explicit discussion of the literature where there has been mixed evidence for the use of the Focus of Expansion. Since the ms is a strong test of the use of curl as a heading signal, this allows a deeper understanding of the importance of the finding and historical context. The ms has also been strengthened by a more explicit discussion of integration of the time-varying signal over periods of several seconds, which is an important demonstration. The implications of the ms are still a little unclear, as the results involve visual judgements in seated subjects. The use of different sources of information when humans walk from one place to another in real life may be complex and involve a variety of different sources of information.

https://doi.org/10.7554/eLife.110770.2.sa1

Author response:

The following is the authors' response to the current reviews.

We thank the editors for their positive assessment of our manuscript, and all the referees for their constructive comments, which have substantially improved this work. We welcome the opportunity to address referee #2's points for the public record, as they highlight key theoretical nuances and valuable future research directions.

(1) We appreciate the reviewer's point that, in a noisy biological system, the increased motion parallax provided by a rich 3D depth structure naturally aids in separating translation from rotation. We fully agree on this point. Our argument aimed at highlighting a fundamental theoretical distinction. Pure algebraic decomposition algorithms are mathematically capable of solving heading on flat planes. The fact that human perception often shows biases in these zero-depth conditions, unless extra-retinal cues are present, suggests that the visual system does not rely on a generalized, global de-rotation algorithm. Instead, it relies on heuristic, depth-dependent structural signals (like motion parallax and retinal curl). We maintain that while depth certainly reduces noise, its strict necessity points toward an ecologically grounded control strategy rather than a noisy global decomposition process.

(2) We think the reviewer raises a valid point regarding the exact neural locus of retinal curl encoding. It is true that Graziano et al. (1994) utilized centred spiral stimuli rather than the spatially offset curl geometries defined in our task. We view the spiral tuning of MSTd not as a direct, one-to-one mapping of full-field retinal curl, but rather as the foundational computational building block required to extract such a signal. We fully agree with the reviewer that identifying exactly where and how this population response is decoded into a unified, gaze-relative retinal curl signal remains an exciting and open empirical question for future neurophysiological research.

(3) We acknowledge the reviewer's call for transparency here. The transition from well-documented multiplicative gain fields (which modulate response amplitude based on eye position) to a direct, localized gaze-centered inhibitory drive is indeed a theoretical abstraction in our model. We utilized this localized inhibition as a functional mechanism to demonstrate how sensory evidence and spatial priors might competitively interact within a standard Mexican-hat recurrent architecture. While gain fields clearly establish that parietal networks integrate gaze position, we agree that the exact local-circuit implementation mapping these gain fields to the specific inhibitory dynamics we modeled has yet to be empirically established.

(4) The reviewer smartly questions whether the 3-5 second behavioral biases delay emerges from the 2.4-second smoothing window used in our computational flow manipulation. It is important to clarify that this 2.4-second window was used solely to stabilize the computed curl signal against high-frequency gait oscillations. This smoothing was restricted strictly to the modeling phase of the controller and neural model and was not applied to the participants' responses. The gradual build-up of their perceptual bias over 3-5 seconds represents their own intrinsic temporal integration of this trajectory, independent of the smoothing parameters used to smooth the curl used in the fitting of the controller and neural modelling.

(5) We concede the reviewer's point that citing a computational model (Layton & Browning, 2014) does not constitute direct empirical evidence for a relationship between MSTd heading preferences and their receptive field locations. Our intention was to highlight a successful theoretical framework that elegantly organizes known properties of MSTd into a system capable of bypassing global de-rotation. We readily acknowledge that direct, single-cell empirical validation of this specific topographic relationship is currently lacking in the literature, and we appreciate the reviewer ensuring this distinction is clearly noted for the record.

The following is the authors' response to the original reviews.

> ***eLife Assessment***
>
> *This study provides an important and biologically plausible account of how human perceptual judgments of heading direction are influenced by a specific pattern of motion*

> *in optic flow fields known as retinal curl. By combining psychophysical experiments and neural modeling, the authors demonstrate that what was previously considered an incidental "nuisance" signal actually serves as a functional control signal for estimating heading and steering toward a fixated target. While the evidence for the role of curl signals is convincing and advances our understanding of vision-based navigation, the work's impact would be strengthened by situating these findings among other cues that contribute to heading estimation, and by clarifying both the time course of these computations and their generalizability across different navigational contexts.*

We thank the editors and reviewers for their insightful feedback and positive assessment of our study. In this revised version, we have made substantial modifications to better situate our findings within the broader landscape of cues contributing to heading estimation, while also clarifying the time course of these computations and their generalizability across different navigational contexts. These points are included in new sections in the revised discussion.

In addition, and following eLife guidelines, we have moved the methods to the end and make sure that the manuscript reads well without needing to go through methods first.

Next, we address all the concerns raised by the reviewers.

> ***Reviewer #1 (Public review):***

We appreciate Reviewer #1's very positive feedback. Incorporating the perspective of 'incidental' sensory signals is a valuable suggestion that aligns perfectly with our findings. We agree that this perspective significantly strengthens the impact of our paper.

In the revised version we have added a last section in the Discussion (Generalizability and Testable Predictions) to comment on the functional utility of 'incidental' signals and incorporated the suggested references. In addition, in the same heading, we briefly elaborate on the predictions and generalizability of the model and possible manipulations that might affect the integration between sensory evidence (curl signal) and straight-ahead prior.

> ***Reviewer #1 (Recommendations for the authors):***
>
> *It would be great if the authors could discuss the implications and predictions of their model.*
>
> *First, from a broader perspective, the study forms an important piece in the emerging recognition that incidental sensory signals are not a nuisance to the sensorimotor system, but contain functionally relevant and effectively used visual signals (Rolfs & Schweitzer, 2022). The authors may want to appreciate their contribution to this perspective in the discussion of the impact of their results. Indeed, a similar shift in perspective has been realized in the recognition that saccade-induced motion signals are not entirely suppressed but play a functional role for gaze correction (Schweitzer & Rolfs, 2021).*
>
> *Second, the authors could spell out additional predictions of their proposal: What are experimental manipulations that could shift the balance between relying on a straight-ahead prior and sensory estimation of curl? What would happen in extreme cases of such sensory evidence? When would priors become overwhelmingly influential?*

As commented in the public review we have now included these two aspects in the discussion.

> *Minor point: After equation 11, the authors may want to specify that, like position, gaze g is also coded as {x,y}, just like position p.*

While the neural model details have been moved to Appendix 2 (including this equation), we added text before Equation 22 (previously Eq. 11) clarifying that gaze is encoded in image coordinates. We omitted point index *i* because the equation applies to all image points relative to a given gaze g.

> ***Reviewer #2 (Public review):***

We appreciate the reviewer's feedback regarding the formalization of our reference frames. We agree that certain definitions were implicitly assumed rather than explicitly stated. We have revised the manuscript to provide all necessary self-contained information, ensuring that the geometry of the task response and the definition of heading are unambiguous. In the last paragraph of the revised introduction, we make clear the response frame of reference which is also included in the caption of fig 1. Also, we have addressed the gap between the task response (in world coordinates) and the functional role of the controller. This is particularly discussed in the discussion (section: reference frames) in which we also provide (and rule out) potential alternatives to our response biases. We also address all the other points raised by the reviewer.

> *Major issues:*
>
> *(1) The manuscript contains inconsistent, if not misleading, messaging about what information retinal curl does, and does not, provide regarding heading estimation. In the Abstract, the authors state: "We propose an alternative: the visual system utilizes retinal curl directly to estimate heading, rendering the explicit recovery of the FOE unnecessary." Based on my understanding of the rest of the manuscript, I find this statement to be a misrepresentation for two main reasons:*
>
> *(a) To "directly estimate heading" relative to what? When not qualified, most people interpret "heading" to mean an observer's heading relative to the world (or some allocentric reference frame). But retinal curl only gives information about an observer's heading relative to the point on which their eyes are fixated. Moreover, that point of fixation will change every few hundred milliseconds in natural viewing, so the retinal curl will change with each new fixation even as heading relative to the world remains unchanged. So I think most readers would grossly misinterpret the claim that retinal curl can be used "directly to estimate heading". Indeed, in the authors' controller model, the initial heading needs to be given, and then the controller can work. But from where does the visual system get the initial heading, since it does not come from curl? These issues are left hanging. Thus, while curl can provide a very useful input for steering toward a fixated target, other signals are needed to estimate heading relative to the world. This has to be made much clearer early on, and a conceptual schematic diagram might help. Also, the authors generally do not specify the reference frame of the variables they are talking about, leaving lots of room for misinterpretations. It should be clear each time they are talking about a variable, such as heading, whether it is relative to the fixation target, body, world, etc.*

In our study, participants were instructed to report their "perceived direction of self-motion" by aligning a rotational encoder (steering wheel) with the direction they felt they were moving within the 3D simulated scene. Consequently, participants reported their instantaneous heading in a world-centered reference frame, from which the 3D trajectories were reconstructed. Since the reviewer had to infer this information, we have clarified this point at the end of the introduction, legend of figure 1, methods (now at the end of the ms.) and discussion to ensure it is immediately evident.

Participants were informed that the initial heading (i.e. θ_0 in our controller nomenclature) was oriented "straight ahead" relative to their body which was aligned longitudinally with

the experimental room. We have modified Figure 1B and revised the Methods section to explicitly clarify this initial alignment and the instructions provided to participants.

In the revised manuscript, we have clarified that while the participant's report is world-centered, the retinal curl provides a gaze-relative heading signal. Although this was already mentioned, we emphasize this point. In natural navigation toward a fixated target, a world-centered vector is often unnecessary; an error signal indicating heading relative to fixation is sufficient (as the reviewer also notes). However, the initial alignment of the heading within the 3D scene allows the brain to "calibrate" this internal controller, mapping the retinal curl signal onto the 3D world coordinates required for the task. Ad commented above, a new section in the discussion addresses and hopefully clarifies the relation with the controller.

The reviewer also asks how we can be certain that participants were reporting in world coordinates rather than an alternative frame, such as "heading relative to the fixation target." We believe our "Cancelled Curl" (and over-cancelled) conditions provide the most compelling evidence to rule out this alternative. In these conditions, the physical position of the fixation target in the scene remained identical to the unaltered flow condition. If participants were simply reporting heading relative to the fixation target's spatial location, the observed biases should have persisted regardless of the flow manipulation. Instead, the bias vanished when the curl was removed. This causal evidence proves that the bias is driven by the retinal motion signal (curl) rather than the spatial orientation of the eyes or the target's position in the scene. Furthermore, the temporal evolution of the response supports a world-centered integration (in agreement with Warren et 2001 Nat Neuro.). For simulated straight paths, the perceived heading remains straight for the first few seconds (consistent with the initial world-centred alignment), with biases only emerging after approximately 3 seconds of integration (a point we elaborate on in our response to Reviewer #3). Had participants been responding based on a simple gaze-relative reference frame from the onset, these biases would have manifested significantly earlier. We have incorporated these points into the revised Discussion to better frame our findings alongside other cues, such as the Focus of Expansion (FOE) and egocentric visual direction that contribute to heading estimation.

Finally, we have rephrased the abstract sentence for clarity. However, we maintain that the original premise remains valid once the world-centered initial heading is aligned with the gaze-centered reference frame.

> *(b) It seems to me that retinal curl will depend on other variables, in addition to heading relative to the fixation target. For example, it seems to me that the magnitude of retinal curl will depend on self-motion speed, the depth structure of the scene, the angle of elevation of the fixated target, and perhaps others. This is not discussed at all, and many readers would get the misguided impression that there is a 1:1 mapping from curl to heading (relative to fixation). If I am right that this is not correct, it means that retinal curl can tell the observer whether to steer right or left to move toward the fixated target, but it cannot tell them how much to steer. Indeed, in the authors' controller model, there is a free parameter that calibrates curl to angle. It makes sense that this works to fit trajectory data that are given from a fixed environment, but it is unclear how the brain would use retinal curl to control steering when these other variables are uncertain or changing unpredictably. Moreover, how does the system change the mapping from curl to steering command as the location of fixation changes relative to the current heading? These are issues that need to be brought up in framing the problem and discussed at some length. If the authors can show mathematically that retinal curl is only dependent on heading (relative to fixation) and not any of these other variables, it would be very valuable to show the equations for this relationship.*

The reviewer notes that we must be clear about the relationship between curl and heading (relative to fixation) and the variables that affect curl. We also thank the reviewer for

encouraging to add the equations that show the relation of curl with additional variables. We have now included these equations in appendix 1.

Beyond the discrepancy between heading (θ) and gaze (ψ), curl is geometrically determined by translational self-motion speed (v), eye height (h), and pitch (α). More specifically, curl = (v.sinψcosα)/h. The derivation is now included in appendix 1. Since h = dsinα, where d is the 3D distance to the fixation point, we could express cos α as a function of distance. Certainly, there is not a 1:1 map from curl signal to heading relative to gaze (e.g. θ-ψ). Participant would need to know v and eye height plus extra-retinal information. Frenz et al (2003, Vis Res.) showed that people can estimate self-motion directly from optic flow, across different simulated eye height and gaze angle; extra-retinal information can, in addition, provide knowledge to ψ and α. It is then plausible that the visual system can use and transform the curl signal from a qualitative directional cue (i.e. steering left or right of fixation) into a quantitative steering command. By combining curl with knowledge of gaze orientation and eye height, the visual system can resolve ambiguities in the flow field and utilize curl as a more precise error signal for locomotor control. These aspects are now included in the new version of the discussion.

> *(2B) I also feel that there is a mismatch between what the behavioral task requires and what the controller model does. Subjects are apparently asked to report their heading relative to the world, but the controller model only controls their heading relative to the point that they are fixating. I understand how this is resolved in the model, but I think this type of distinction is buried and will not be apparent to most readers. Again, the reference frames of what is being measured and controlled need to be specified explicitly in all parts of the paper, and the authors need to explain how the system would combine curl-based control with some other measures of (at least initial) heading for world-centered heading to be computed. All of the assumptions need to be clearly specified.*

We thank the reviewer for this point. We have addressed the alignment of the reference frames in our response to Issues 1a and 2a. Once the initial orientation (θ_0) is established in the world frame, the controller model generates steering adjustments that directly translate into heading predictions within that same world reference frame. By treating the perceptual report as an output of the locomotor controller, we resolve the discrepancy between the steering task and the reported heading.

> *(2c) In addition, I found it frustrating that the authors never present raw perceptual data from the observers. Rather, in Figure 2, we see reconstructed trajectories that are perfectly smooth with no indications of noise whatsoever. Since these paths are computed from the perceptual reports, there must be some noise inherent in them. The figures should represent this uncertainty somehow, and it should be explained how these perfectly smooth trajectories are obtained.*

We respectfully disagree with the reviewer's interpretation regarding data smoothing. The thin lines in Figure 2 represent the mean 3D paths derived directly from the response variable (θ_t) across trials of identical conditions for each participant (as detailed in the 'Computation of Perceived Path' section). No smoothing or filtering has been applied to these plotted trajectories other than computing the mean across trials. We also wish to remind the reviewer that the raw data and analysis code remain publicly accessible for further inspection. Having said that, we include now a supplementary figure showing an example of raw data responses as a function of time. This figure will be a supplemental figure of main Figure 2 (now provisionally included in the Suppl Information).

Regarding the visual representation: in earlier versions of the manuscript, we included shaded 95% Confidence Intervals (CIs) in Figure 2. However, this addition rendered the plot overly cluttered and obscured the individual trajectories. We therefore chose to present individual participant means (thin lines) alongside group averages (thick lines) to emphasize

inter-subject variability. For clarity, the 95% CIs are explicitly displayed in Figure 3, where the data density is more conducive to shaded areas.

> *(3) "...the magnitude of retinal curl in the fovea can specify the body trajectory relative to gaze (Matthis et al., 2022)." The main idea put forward by the authors here seems to overlap heavily with this statement that they attribute to Matthis et al. 2022. While I think this paper still adds importantly to the topic, the authors do not discuss how their findings are different from those of Matthis et al. 2022, why they are an important extension, etc. Readers should not have to go read this other paper to have any idea how the present findings are placed in importance relative to the literature.*

We have updated the Discussion to more specifically align our findings with Matthis et al. (2022). We emphasize that our study provides the perceptual validation for their ecological observation that the FOE is often too unstable for reliable use, whereas foveal curl remains a robust signal for path estimation. Our paper provides the causal link, since we manipulate curl in real-time (the 'cancelled & over cancelled curl' condition) providing the critical evidence that perceived heading is affected by this signal. The relation with this previous study is made clear in the revised discussion.

> *(4) The analysis and treatment of eye movements is extremely weak. The authors discarded trials for which gaze deviated from the fixation point by more than 3 degrees (which is a LOT given that the eye speeds are generally in the neighborhood of 0.5 deg/sec), and they provide basic stats on the distribution of positions. But this largely misses the point: it is not small position errors that are likely to matter, but rather velocity errors. Even a small amount of retinal slip of the target while it is being pursued will cause image motion that is going to alter the optic flow field around the fixation target. So, for example, the retinal curl field may no longer be centered on the fixation target. How do we know that some of the perceptual biases are not influenced by image motion resulting from imperfect tracking of the fixation target? This needs to be analyzed and discussed.*

We thank the reviewer for noting that retinal slip (velocity error) is a more critical metric than positional gaze error. We agree that tracking inaccuracies can introduce translational noise into the flow field. The 3° threshold was established based on the eye tracker's specifications and the naturalistic setup (1-meter viewing distance without head stabilization). Across all participants, the mean positional error ranged from 1.016° to 1.5° (1 deg is 2.08 cm in our setup). We also calculated retinal slip values, which ranged from 0.12 to 0.27 deg/s (X dimension) and 0.12 to 0.23 deg/s (Y dimension). These values are comparable to natural oculomotor drift (Kowler et al., 1979) and are understandably small given the low velocity of the fixation target. We have added this information about retinal sleep at the beginning of the results section.

Consequently, it is highly unlikely that retinal slip influenced the results. Furthermore, assuming that tracking error remained consistent across fixation conditions, any present retinal slip cannot explain why the bias followed the retinal curl manipulation as predicted by the controller. We therefore consider retinal slip to be an unlikely confounding factor.

> *(5) I found the sections of text comparing the separate and joined fits (starting line 287) to be a bit too rosy. The authors show the separate fits in the main text, and it is not very surprising that these fits are good, given that the model has 30 parameters, and these data are pretty low-dimensional. The authors only show the joined fits in the supplement, and they say that they are almost as good as the separate fits (indeed, they are better in a model comparison sense, but this is 30 parameters vs. 2 parameters). However, when I look at the fits of the joined model in the supplement, I don't find them to be very impressive. In particular, the model grossly misses the data for the straight paths for several subjects (e.g., id5, id6, id8, id10). And fitting the straight paths would presumably*

> *be easiest. This implies that the joined model is really missing something and that fitting the curved paths interacts strongly with fitting the data for different fixation target locations on the straight path. I think that the authors should discuss the results a bit more soberly and tone down their conclusions here.*

We thank the reviewer for the opportunity to clarify the logic behind our modeling choices. We acknowledge that the "separate fits" are inherently less informative due to the high number of free parameters relative to the data. Our primary scientific goal was not to achieve perfect descriptive accuracy via 30 parameters, but to test a specific functional hypothesis through the "joint fit."

The Logic of the Joint Fit:

We agree with the reviewer that the joint fit misses some paths in some conditions. Of course, the joint fit reflects a significant compromise. The "Gain" (the weighting of the curl signal) is likely not a static constant but is dynamically tuned based on task demands, confidence in the visual signal, simulated speed, and so on. By using a single Gain parameter, we intentionally ignore this contextual variability to see how much of the behavior can be explained by a "minimalist" controller. In this sense, the 2-parameter joint model is a deliberate attempt to test this limit. By forcing a single Gain parameter to account for all conditions across both straight and curved paths within one flow manipulation (e.g. unaltered flow) we are asking if a single, fixed linear relationship between retinal curl and steering effort/gain can explain the results. We view the joint fit not as a "perfect" model, but as a stronger test of the curl-based control theory. The fact that a 2-parameter model can capture the direction and scale of biases across such a diverse set of conditions (straight/curved paths, five fixation eccentricities) suggests that retinal curl is a robust signal. Upon closer analysis, these discrepancies between the joint model and the data are most pronounced in the over-cancelled condition which is the one when sensory evidence becomes more ecologically inconsistent with the extra-retinal information (gaze direction). While the joint fit successfully demonstrates that a single parameter can capture the general functional role of curl, it fails to account for the complex sensory re-weighting that occurs in ecologically inconsistent conditions (like 'over-cancelled' flow). We have updated the manuscript to discuss these limitations in the "fitting the controller" section, framing the model as a parsimonious first-order approximation rather than a complete description of human heading perception based on a minimal set of parameters.

> *(6) The section of the paper on neural simulations (starting line 387) has a few weaknesses. First, why are only straight paths simulated here? This does not seem to provide a very rigorous test of the model. Second, it is awkward that the simulation results are presented in units of pixels, rather than degrees. Third, the authors seem to downplay the fact that the neural estimates of heading seem to oscillate rather wildly (over a range of hundreds of pixels, whatever that means, see especially Figure S16). It was far from clear to me how an estimate of heading with these large oscillations is useful. It would seem to require that heading estimates are integrated over substantial lengths of time to be reliable. It was therefore unclear how the model produces such smooth paths from these oscillating estimates.*

We acknowledge that the presentation of the neural model requires more clarity regarding its objectives and its relationship to the behavioral data.

We first wish to clarify the intended scope of the neural ring-attractor model. Our primary goal was not to provide a comprehensive account of behavioral performance across all conditions (which is the role of the controller model), but rather to demonstrate a biologically plausible mechanism that explains the emergence of the "Opposite-to-Gaze" bias. While the controller demonstrates that the bias follows a specific control law, the neural model shows

how such a law can emerge from known primate neurophysiology, specifically, spiral-tuned MSTd neurons, gaze-contingent inhibition, and an egocentric “straight-ahead” prior.

Why Straight Paths are Sufficient for this Objective. The reviewer asks why only straight paths were simulated. In our study, the straight-path condition with eccentric gaze is the purest test of the bias mechanism. Simulating the straight paths allowed us to isolate the interaction between foveal inhibition and the straight-ahead prior without the confounding variable of path-curvature flow. Given the complexity of the neural network’s parameter space, we focused on these conditions to provide a clear neuro-plausible explanation. We have added text when introducing the model (Neural simulations in the Results section) to make clear why we model straight paths only.

Units: Pixels vs. Degrees. We acknowledge that the use of “pixels” in the plots of internal neural dynamics may appear awkward. The neural network operates on input stimuli that are defined by the pixel resolution of the videos used in the simulations, we used pixels as the native coordinate system to describe the movement of activity peaks within the network’s internal “map.” We have decided to keep the pixel units in these figures.

Behavioral Output (Meters): Importantly, the final heading estimates produced by the network are not left in pixels. We use a pinhole camera model to reconstruct the 3D trajectories from the neural activity. These results are expressed in meters, allowing for a direct comparison with the human behavioral data.

Addressing Wild Oscillations and Smooth Paths. The oscillations observed in the instantaneous heading estimates reflect the stochastic nature of the population peak when tracking high-frequency sensory inputs. In our model, the synaptic time constant (τ) was kept relatively small to ensure a fast, low-latency response to changes in self-motion. While increasing τ would have produced smoother internal dynamics, it would also have introduced delays into the control loop. Instead, we chose to maintain this high sensory responsiveness and applied a temporal moving average later to the network’s decoding to reconstruct the 3D trajectories. This is explicitly stated in the section “Heading Estimation and 3D path reconstruction” in the new appendix 2.

In addition, the neural activity over time is shown in two ways: the heatmap shows the neuron with preferred heading (one can see more oscillations, specially when the fixation point is closer to the centre (eccentricities -2 and 2), due to larger competition between the sensory evidence and the straight-ahead prior. The other way is the decoded heading. In the ring-attractor model, the decoded heading (ϕ) is not determined by a single neuron but is calculated using a population vector average (equation 19). By summing across the entire population, the decoder effectively integrates sensory evidence from many neurons simultaneously. One can appreciate (see e.g. Fig. 5B) that averaged decoding, leads to a smoother resulting estimate (the white dashed line, whose visibility had been improved in the revised version). Behavioral work by Burr and Santoro (2001) suggests that global motion signals (divergence and rotation in optic flow) are integrated over much longer timescales—roughly 1000ms to 3000ms—compared to local motion units (~200 ms).

In the previous manuscript, we discussed this aspect in lines 424-426. In the new version, we have added text in the Heading estimation and 3D path reconstruction section (now in appendix 2) stating that we smoothed the decoded signal in agreement with this psychophysical evidence before applying the camera model.

See also our comment on temporal integration in the responses to reviewer #3

> ***Reviewer #2 (Recommendations for the authors):***
>
> *(7) Line 51: "...a functional role of rotational flow components has been largely neglected in both theoretical and experimental work on heading perception." I feel like this*

> *statement is too strong and that the authors try too hard to "sell" their findings by underrepresenting previous work. There are numerous studies (many not cited), both behavioral and electrophysiological, that have examined how heading perception depends on pursuit eye movements, either physical movements or visually simulated ones. These studies directly involve rotational flow components, and several of them have concluded that rotational flow components contribute to estimating heading in the presence of eye movements (just one example is Grigo and Lappe 1999). Because these studies generally involved horizontal pursuit of a target on the horizon, rather than tracking a point in the ground plane (like the authors' work), these studies generally did not involve flow fields with retinal curl around the fixation point. But I consider these older studies just a special case of the more general geometry, and they still involve rotational flow components. Moreover, various previous studies have used stimuli for which there was no FOE present in the visible display (either due to simulated rotation or masking out the FOE), and the authors do not seem to give credit to these works either. In addition, several studies have implicated a role of extraretinal signals in perceiving heading during eye movements, so retinal curl cannot explain everything. Rather than overemphasizing the limitations of previous work, the authors would be better served to explain how their findings extend and generalize from these previous studies.*

We thank the reviewer for pointing out this oversight; it was not our intention to overlook previous work. While our original version cited studies considering rotation-related cue, we have now substantially revised the introduction to include previous work and better acknowledge the informative role of rotation. Our central aim remains to distinguish between models that compensate for rotation to recover a heading vector and our proposal that the visual system exploits retinal curl directly as a primary, functional signal for locomotor control.

We have now updated the Introduction and Discussion to better situate our work within the context of studies (including Grigo & Lappe, 1999 and some additional ones we have included in the new version) that have investigated the informative role of rotational flow. We now clarify that our study extends these findings by investigating the non-uniform rotational patterns (curl) that emerge during ground-plane fixation, representing a more general and biologically ubiquitous case of locomotor control, while acknowledging previous studies that have also considered the potential role of curl generated by gaze fixation.

> *(8) Figure 3: I did not understand why there are two purple and two blue curves in the graphs of the middle column. And the caption does not explain this.*

This a very good observation. This was explained in lines 268-273 (previous version). When the gaze is straight-ahead (same direction as heading), there is no curl. However, we introduced positive or negative curl in the altered conditions. These purple and blue lines refer to these trials and show that the bias re-appears in the expected direction when curl is (unexpectedly) added. We think this adds additional evidence to the curl contributing to heading. Even though this was extensively explained we have added text in the caption of figure 3.

> *(9) Line 331: What makes the authors think that retinal curl is computed in area MSTd? They should cite studies to support this idea if it has been shown in physiology.*

Evidence was cited in the introduction (Graziano et al 1994) of sensitivity to spiral motion in addition to neuro-computational models that implement this activity also cited (e.g. work of Leyton et al.)

> *(10) The neural network model for computing heading from curl requires a "gaze-centered inhibitory drive" that inhibits activity around where the eyes are looking. This is probably a biologically plausible thing, but is there any evidence to support the idea that*

> *this signal exists in the parts of the brain where the authors believe these computations to be happening? They simply posit the existence of this gaze-centered inhibition as though it is common knowledge, but they provide no citations nor discuss any previous evidence for its existence.*

While neurophysiological evidence primarily describes this as gain-field modulation, this process frequently involves localized suppression of neural activity to facilitate coordinate transformations. In parietal areas such as LIP and 7a, eye-position signals do not just enhance responses but can also suppress them, effectively shifting the 'center of gravity' of a population response (Read et al 1997; Born et al 2005, cited in the discussion in the revised section re-evaluating the Focus of Expansion). In the context of our ring-attractor model, this functional modulation is most parsimoniously implemented as a gaze-centered inhibitory drive.

> *(11) Lines 482-483: Why should perceptual biases related to retinal curl take seconds to show up?? The curl information itself must be present very quickly, perhaps requiring just a few video frames. So what does this imply about mechanisms? The authors throw out this assertion, but it is left hanging without any further analysis or support.*

The time course reflects the integration requirements of complex motion processing. While local flow is processed rapidly, global patterns like retinal curl require longer temporal windows to reach a stable estimate (Burr et al 2001). In our study, this integration is functionally necessary to filter the higher-frequency 'wobble' induced by gait-cycle oscillations. We now discuss the temporal integration aspects under a new heading in the discussion.

> *(12) Line 508: "This suggests that the "bias" observed in our perceived headings may reflect the operation of a control law optimized for action rather than a failure of a perceptual system designed for passive estimation." The authors make this statement to justify why perceptual biases are present with unaltered curl. But I don't fully understand the logic. Are they saying that it is not possible to have a set of computations that can do both things accurately? Is it possible to show this theoretically? Moreover, if it is not possible to rule out other possible sources of the biases, such as those described above (reference frame of judgments, eye movements, etc), then is it necessary to invoke this logic?*

Our logic is that the observed 'bias' is not a representational failure, but a functional byproduct of a control law optimized for active steering. In a closed-loop system, the objective is to null the error signal (retinal curl) to maintain a stable path. When observers are asked to make an open-loop heading report, they likely utilize this same control signal, which manifests as a systematic bias toward the 'null' point of the controller as a result of a sustained fixation in discrepancy with the simulated translation/heading.

We do not suggest that accurate perception and control are theoretically incompatible; rather, we suggest that perception and action rely in the same underlying information (e.g. work of Brenner & Smeets). While other factors, such as coordinate transformations between reference frames, certainly can contribute to the reporting process, our interpretation provides a parsimonious link between the psychophysical data and the underlying steering mechanism. By framing the bias as a consequence of a 'nulling' strategy, we explain not just the existence of the error, but its specific direction and magnitude relative to the fixated target.

> *(13) Line 546: "...MSTd would simultaneously code curvature for trajectory estimation and heading across the neural population, with curvature encoded through the spirality of the most active cell and heading through the visuotopic location of its receptive field center." The latter part of this argument seems to imply a relationship between the*

> *heading preferences of MSTd neurons and the locations of their receptive fields. I am not aware of any evidence for such a relationship, so the authors should indicate whether this is based on some experimental data or just a speculation.*

We thank the reviewer for this observation. The proposal that heading is signaled by the visuotopic location of active MSTd populations is a core architectural feature of our model and is supported by several lines of evidence. In the Layton and Browning (2014) framework, MSTd is modeled as a visuotopic map of functional 'hypercolumns'. Each hypercolumn contains neurons tuned to a continuum of spiral patterns, but all neurons in a given hypercolumn share a receptive field center at a specific location in visual space. Consequently, the visuotopic location (x, y coordinates) of the maximally active hypercolumn represents the center of motion (heading), while the spirality (the tuning dimension within that hypercolumn) represents path curvature. We have clarified this in the revised discussion (re-evaluating the FoE) to emphasize that this dual-coding scheme arises from the simultaneous representation of 'where' (population map location) and 'what' (spiral tuning) in MSTd.

> ***Reviewer #3 (Public review):***
>
> *The primary limitation of the paper is that it avoids discussion of some of the inevitable complexities of heading perception. The main issue is what exactly is meant by heading. Different behaviors evolve over different timescales. The geometry of retinal motion defines instantaneous heading, which varies widely through the gait cycle. Time-varying information like this is known to be important in the momentary control of balance. Heading can also be thought of as steering the body toward a distant goal, which evolves over longer timescales. The current manuscript appears to be concerned with heading information integrated over a few seconds and seems to provide evidence that heading is indeed integrated over the gait cycle. The issue of the time scale of the computation is touched on, but it is not related to how it might be used in normal walking or what situations it might apply to. Steering toward a distant goal during walking is not a very difficult problem and may not require evaluation of retinal motion, but control of balance is more challenging and may depend critically on curl. Consequently, the timescale of the computation needs to be considered in order to understand what is meant by heading.*

We thank Reviewer #3 the comments regarding the definition of heading at different time scales, the role of the gait cycle, and the temporal integration of the curl signal. These comments have helped us refine the manuscript's core arguments.

We agree that "heading" must be precisely defined within the context of the differing temporal demands of balance and steering. While instantaneous heading provides the high-frequency feedback necessary for momentary postural adjustments and balance, our study is concerned with heading as a gaze-relative signal used for the continuous control of a locomotor trajectory. As such, we have revised the manuscript to specify that the perceived heading measured in our task reflects a signal integrated over the gait cycle to filter out the oscillatory noise induced by head bob and sway (mainly in the Discussion section).

The reviewer correctly notes that gait-induced head bob and sway produce high-frequency oscillations in the curl signal, yet our behavioral results show smooth, slowly evolving biases. The visual system does not react to "instantaneous" curl, which would lead to jittery, unstable heading estimates. Instead, it integrates flow over a timescale roughly commensurate with a full gait cycle (~500–1000ms). This implies a significant temporal integration process. This temporal integration is consistent with evidence (Burr and Santoro,2001, Vis Res) indicating that optic flow signals (radial and rotational components) are integrated over windows of approximately up to 3 seconds to ensure perceptual stability. Neurally, this likely involves the projection from area MSTd to the Ventral Intraparietal area (VIP), a pathway where fast, eye-

centered sensory inputs are transformed into stable, body-centered representations suitable for guiding long-term steering behavior (Chen et al. 2011, JNeurosci.). By grounding our definition of heading in these specific temporal and neural constraints, we tried to clarify how the visual system exploits retinal curl for goal-directed action in natural, dynamic environments and relate our findings to recent studies addressing the role of retinal motion on balance (Powell et al. 2026 Bioarx).

In our implementation, we explicitly address the high-frequency noise introduced by gait dynamics by smoothing the retinal curl signals computed from the stimulus videos before they are fed into the controller. This temporal filtering allows the fit of the controller's prediction to the response data while remaining robust to the rapid fluctuations of head bob and sway. In contrast, the neural ring-attractor model would not require an external smoothing step; instead, the integration is an emergent property of the system's architecture that can be controlled with different parameters, as commented above in a response to Reviewer #2. The dynamics of the synaptic weights and the characteristic "leak" in the population activity naturally implement a leaky integration of sensory evidence, ensuring that the decoded heading reflects a sustained estimate rather than an instantaneous response to visual noise.

We also agree that we avoided discussing some complexities of the heading perception. In the new version, we also include and integrate the distinction between instant heading and future path in different parts of the ms (introduction) and mainly discussion (temporal integration and steering section) which have been revised substantially.

> ***Reviewer #3 (Recommendations for the authors):***
>
> *There are a number of points that require clarification.*
>
> *(1) Head bob and sway were included in the stimulus and need to be addressed in both the analysis of the data and the interpretation. The curl signal in the stimulus varied over time, commensurate with normal gait. However, the results don't reflect the same level of variability that would be produced from curl over the gait cycle. This means that the information must be integrated over some longer timescale. It is not clear from the data analysis what this integration is. Is there an implicit integration with the manipulation of the steering wheel? If subjects indeed appear to be able to use curl to evaluate heading over timescales of seconds, this needs to be explicitly addressed, as it is a novel result. This would require parts of the discussion to be changed/expanded to maintain consistency. For example, line 482 talks about the buildup of biases over time.*

As commented above in the public response, the curl estimated from the optic flow algorithm was smoothed before being input into the controller (path fitting and predictions). The smoothing was only applied to the curl signal, not to the participants responses. Also, as mentioned before, the time course of the bias is consistent with integration times of optic flow reported in the literature. All these aspects are now explicitly included in the new display and conditions section (Flow manipulation conditions).

> *(2) Since the experiment included curl variability resulting from the gait cycle, some discussion is needed about the role of retinal motion in the control of balance and posture. There is a large literature about the role of flow in controlling gait and momentary adjustments of the body while walking. Additionally, it should be noted that in the task, head bob and sway from 1 prerecorded individual was shown to all subjects. It is known that gait varies significantly between different individuals, and it should be acknowledged that this may lead to differences at the individual subject level for perceiving heading.*

We have included a point in the discussion addressing the different time scales for different use of optic flow signals (postural control vs locomotion).

We agree with the reviewer that utilizing a single gait profile for all participants may introduce individual differences in perceived heading, as the simulated head motion might not perfectly match each participant's unique biological gait signature. However, we prioritized stimulus consistency over idiosyncratic accuracy. By ensuring that every participant viewed the exact same motion profile, we could be certain that the systematic 'opposite-gaze' biases observed across the population were driven by our experimental manipulations of gaze and retinal curl, rather than being confounded by variability in head-motion kinematics. We have added an acknowledgement of this point at the first paragraph of the displays and conditions section in the Methods.

> *(3) More information is required about the use of the rotating wheel for the measurement of heading. How easy was it to use? What about time delay, and how does this deal with the bob and sway? Does the wheel impose a de facto integration on the perceptual measurement?*

The rotary encoder provided an intuitive steering-wheel interface that participants found easy to operate. To ensure minimal latency (1–5 ms), the device was interfaced via an Arduino Uno and sampled by a dedicated background Python thread, isolated from the visual rendering loop. We have incorporated these technical details into the Methods (Procedure) section.

> *(4) Restructuring the description of the models It remains unclear why the dynamics of the neural network are a necessary inclusion in this paper. It seems interesting, but there is no comparison to actual neural data or other related work. Instead, this appears to be a description of what the network is doing, which is already defined by the equations. This needs to be clarified for its exact interpretation with respect to real neural data, and its importance here for understanding the biases that emerge in heading judgments. The paper would flow better if this section were included as supplementary material or omitted from the paper entirely, as it seems to detract from the other points. If this is a description of why the biases are seen in the controller, then the supplementary material is a good place for it.*

We thank the reviewer for this suggestion. We clarify that the neural model is not intended to simulate specific empirical neural data, but rather to provide a biologically plausible implementation of the controller. This allows us to demonstrate how the observed biases emerge from the dynamics of standard cortical architectures, such as ring attractors. This is now mentioned when introducing the neural simulation results.

Following the reviewer's suggestion, we have moved the neural model equations to Appendix 2 while retaining the simulation results in the main text (Results). We believe it is essential to present not just the abstract controller, but also its functional implementation, as this provides a mechanistic bridge between retinal signals and locomotor behavior.

> *(5) For the modeling approaches, the math would be more appropriate for supplementary materials.*

To ensure a better flow of the paper, we have moved the neural model equations to Appendix 2, while Appendix 1 now details the relationship between measured curl and other variables (speed, yaw, pitch, etc.). We have retained the controller model in the Methods section, consistent with the eLife layout where Methods follows the Discussion.

> *(6) How do the models and data analysis deal with the influence of the gait cycle in the input? Do they integrate the information over that timescale? If so, the integration time*

> *needs to be specified.*

As commented above, to mitigate gait-cycle fluctuations, we smoothed the computed curl signal before inputting it into the controller and applied a similar smoothing process to the neural model's readout. Using a LOESS filter, the effective integration window was 2.4 seconds. Close to the integration time reported in Burr et al. 2001. These parameters have now been explicitly specified in the Methods section. For the empirical data, we just utilized trial-averaging.

> *(7) What does biologically plausible mean in terms of the neural network model? Especially when control wasn't explicitly a variable in the measured behavior of the subjects.*

By biologically plausible, we mean that our model is constrained by neural architectures documented in the primate brain—specifically ring-attractor dynamics, population coding, and gaze-centered gain-fields. Crucially, the network utilizes recurrent connectivity with a 'Mexican-hat' profile (local excitation combined with lateral inhibition). This is a standard and widely accepted motif in computational neuroscience, representing the consensus on how cortical circuits maintain a stable "bump" of activity to represent spatial variables. Rather than introducing ad-hoc mechanisms, we demonstrate that the observed behavioral biases emerge naturally from these established neural components when they are tasked with maintaining locomotor stability. Since we think this aspect was already emphasized, we haven't added any additional detail.

> *(8) There should be more extensive acknowledgement of the body of literature that has challenged the use of the focus of expansion. That section should also include references to work that has investigated extraretinal signals, as they may also be important.*

We have expanded the Introduction and Discussion to more thoroughly acknowledge research challenging FOE-based models and the critical role of extraretinal signals. These updates, which also align with our response to Reviewer #2, provide a more comprehensive context for our model within the existing body of heading and self-motion literature.

> *Minor points:*
>
> *(1) Line 69 - In self-generated motion, spiral patterns are almost always centered on the fovea, but many physiological experiments present spirals in the peripheral retina. This is incompatible with the motion generated during self-motion. Therefore, clarify whether the type of spiral motion Graziano investigated was centered on the fovea.*

In the experiments conducted by Graziano et al. (1994), spiral stimuli were centered on the receptive field (RF) of the individual neuron being recorded to accurately characterize its tuning. While the reviewer correctly notes that spiral centers often align with the fovea during active steering (due to fixation on a goal), MSTd neurons possess large RFs that provide a comprehensive 'template' system across the visual field. This population-level representation allows the brain to recover trajectory information even when the focus of motion shifts relative to the fovea—for example, during pursuit eye movements or when fixating on landmarks off the direct path of travel. To keep this part of the text brief, we haven't add more details concerning this study.

> *(2) Line 72 - "Magnitude" instead of "amount".*

This has been changed.

> *(3) Line 115 - State explicitly whether the scale of the visual stimulus was matched to the scale of the actual natural images shown in VR.*

We have updated the Methods (Displays and conditions) section to explicitly state that the visual scale was veridical. The virtual camera's parameters were calibrated such that its field of view (91°) matched the physical dimensions of the projection screen (2.03 m × 1.16 m) at the 1.0 m viewing distance. This ensures that the angular size of the objects and motion gradients in the stimulus were 1:1 with the scale of the simulated natural environment.

> *(4) Line 144 - While Farneback is a good dense flow estimation algorithm, it is noisy and may impose biases/variability in the calculation of curl. This should be acknowledged.*

We acknowledge that the Farneback algorithm can introduce variability in curl estimation. To mitigate this, we utilized 10 independent renderings of each experimental trial to compute the flow fields. Although this methodology was reflected in the data previously uploaded to our OSF repository, we have now explicitly added this detail to the manuscript (Flow (curl) manipulation conditions). The computed curl used for the modeling was derived from the aggregate of these different runs, ensuring a robust and stable signal that accounts for potential algorithmic noise.

> *(5) Line 206 - "Join fits". Is this a technical term? It sounds awkward. Would "Joint fits" make more sense?*

The referee is right. We have corrected this.

> *(6) Line 280 - "Consistent with" (typo).*

This has been corrected.

> *(7) Line 286 - Typo in title.*

Also corrected to Fitting the controller.

> *(8) Lines 482-493 - There should be more discussion on the time course of integrating the stimulus, and the relationship/generalizability to more natural stimuli.*

This part of the discussion (related to integration time) has been changed considerably to include discussion of postural control in addition to locomotion.

> *(9) Lines 516-517 - It is mentioned that retinal flow dynamics override the visual direction cue. This may not be generally true, as the reweighting of the cues might depend on things like task demands or actual stimulus context. In the present experiment, the subject only has access to a large moving textured ground plane, and the body is stationary.*

We agree with the reviewer that cue reweighting is highly context-dependent. However, as noted in the original manuscript (Lines 516-517), we specifically stated that retinal flow dynamics 'can' override the visual direction cue, rather than asserting a universal rule. This phrasing was intentional to acknowledge that while flow is a potent signal—especially in the presence of a large, textured ground plane as used in our paradigm—the relative weighting of these cues remains contingent on the specific sensory and task conditions. We believe this remains a fair and cautious interpretation of our findings.

> *(10) Lines 524-256 - It is unclear why Matthis et. al. 2022 is cited for this point. Some of the steering literature, like Wilkie Wann & Allison 2006 or Lappi & Mole 2018 (and some of their other work), would be more relevant for the definition of a control law under these circumstances.*

We agree and this part has been changed substantially.

> *(11) Lines 528-532 - Warren et. al. 2001 should be cited in this section because their results were interpreted in terms of focus of expansion, but may result from the curl signal (Powell et. al., 2026. The Role of Retinal Flow in Walking. bioRxiv, 2026-02.).*

We agree with this suggestion and the citation has been added.

> *(12) Lines 544-546 - Layton and Browning are focused more on path perception from the implemented spiral tuned cells. Because of this, it wouldn't be appropriate to say it is a shift away from FOE-based heading based on this citation alone. There are more models/psychophysical results that would strengthen this claim.*

We agree with the reviewer that Layton and Browning focus specifically on path perception. Our original intention in citing this work was to emphasize the neurophysiological continuum from radial to circular motion (spiral tuning) rather than discrete expansion/rotation channels. However, we have revised this section to clarify that the reliance on retinal curl represents a mechanism for determining the future path (locomotor trajectory) rather than merely instantaneous heading. This distinction acknowledges that while heading is a momentary vector, the integration of curl signals allows the system to anticipate and control the intended path over time—a framing that better aligns with both the cited literature and our proposed controller model. As commented above this is now extensively discussed in the revised version.

> *(13) Figures - There are minor visibility issues for some of the figures. In Figure 2, the thick line is unreadable, and in Figure 1, the x-axis labels are crowded.*

The axis in Fig 1 has been modified to avoid crowdedness. In Fig. 2, the thick (average line) has been modified. We hope they are more visible now.

https://doi.org/10.7554/eLife.110770.2.sa0