Flexible control of behavioral variability mediated by an internal representation of head direction

Chuntao Dan[1], Ramya Kappagantula[1], Brad K. Hulse[1],
Vivek Jayaraman[1*], and Ann M. Hermundstad[1*]

[1]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

*correspondence to: vivek@janelia.hhmi.org, hermundstada@janelia.hhmi.org

ABSTRACT

Internal representations enable flexible behavior in both animals and artificial agents. Here, we propose a conceptual framework for how *Drosophila* use their internal representation of head direction to maintain and modify preferred headings upon selective thermal reinforcement. We show that flies in a well-established operant visual learning paradigm use stochastically generated fixations and directed saccades to express heading preferences, and that silencing their head direction representation compromises their ability to modify these preferences based on reinforcement. We describe how flies' ability to quickly adapt their behavior to the rules of their environment may rest on a behavioral policy whose parameters are flexible but whose form and dependence on the head-direction representation are genetically encoded in the structure of their circuits. Many of the mechanisms we outline may be broadly relevant for rapidly adaptive behavior driven by internal representations.

INTRODUCTION

Behavior often depends on the transformation of sensory information into motor commands based on an animal's internal needs. Some direct responses to sensory stimuli do not require a brain [1] or even neurons [2], but neural networks enable animals to more precisely direct their actions, and to adapt their responses to sensory stimuli based on context, internal state, and experience [3–5]. However, sensory cues are not always reliable or even available, and many animals have evolved the ability to behave still more flexibly by generating and using internal representations of their relationship to their surroundings [6, 7]. These internal representations —for example, those carried by head direction (HD), grid, and place cells [8, 9]— are often tethered to sensory cues, but they allow animals to achieve behavioral goals without directly depending on those cues. Thus, goal-oriented behavior can sometimes operate in two phases: first, a latent learning phase in which an animal explores and acquires an internal model of the structure of its environment, and second, a phase in which the animal modifies its behavior in that environment through its good or bad experiences in specific situations and/or places [10]. Many studies of learned behavior and its neural correlates, particularly those involving mammals, focus on the second phase, using trained animals that have already learned the basic structure of tasks and environments; in doing so, they study task performance more than task acquisition (but see, for example, [11–15]). By contrast, in many natural settings, animals must develop internal representations of their spatial relationship to their surroundings at the same time that they discover the environment's rules through reinforcement of specific actions or contexts. Further, they must use these still-evolving representations to select appropriate actions (or, in the parlance of reinforcement learning (RL) [16], to guide their behavioral policy). Of course, even in natural settings, animals do not already know the exact tasks that they will be required to perform. Rather, they must often infer the specifics of their behavioral goals based on the consequences of their actions. Moreover, behavioral goals can themselves change over time based on environmental conditions and internal state, requiring animals to balance exploitation of a good situation with explorations away from it.

In this study, we delve into the dynamic process by which internal representations, goals, and behavioral policies develop and work together in the context of rapid, visually-guided operant learning in the fly, *Drosophila melanogaster* [17]. We used a variant of a well-established learning paradigm for tethered flies [18] to explore the behavioral policy —and the underlying circuit architecture— that enables flies to modify their actions in response to heat punishment associated with one of two repeating visual patterns arranged symmetrically around the fly (Fig 1a). We found that silencing so-called EPG or compass neurons that maintain a representation of flies' head direction (HD) [19] significantly impairs performance in this paradigm. We quantitatively analyzed the behavior of individual tethered

flies to understand the structure of their actions, used this analysis to isolate a set of control parameters that govern the action selection process, and showed how these control parameters should optimally be structured in order to maintain a behavioral preference. We showed that flies' behavior is consistent with such a heading-dependent policy, but only if the symmetry of the setting evokes a predictable instability in the internal HD representation—equivalent to the fly's 'internal heading' in this setting and therefore used interchangeably here (see Supplemental Information for further discussion of this issue). Recent studies have shown that the fly's internal representation of HD is observable as a 'bump' of calcium activity in compass neurons in the ellipsoid body (EB) [19], a substructure of an insect brain region called the central complex (CX) [20–26]. We used two-photon calcium imaging of compass neurons in tethered flying flies to examine the HD representation in the visual scene used in our behavioral paradigm. We found that the symmetry of the visual setting indeed induces a structured instability in the fly's HD representation. We used these observations, together with existing physiological, behavioral, and anatomical observations [19, 27–34], to construct a model for how CX circuits downstream of compass neurons might use the HD representation to both learn a goal heading in its visual environment and select actions that are driven by the fly's current heading relative to its goal. The fly's behavioral policy appears to be tethered to the difference between the fly's current and goal-heading and its *form* seems hardwired; not needing to acquire the behavioral policy from scratch for each goal heading likely enables rapid learning and immediate adjustments to new goal headings. Finally, we used the model to understand the interplay of the processes governing the generation of the HD representation in this environment, the inference of the goal heading, and the selection of actions. Our results and conceptual framework cast decades of influential fly visual learning studies [18, 27, 35, 36] in a new light, suggesting that learning in these tasks operates on an internal HD representation rather than directly associating actions with reinforced visual patterns. These results also provide a window into how rapid behavioral flexibility can be enabled by a pre-structured yet adaptive policy tethered to an internal representation, and they underscore the importance of the reliability of that representation in shaping the animal's behavior.

RESULTS

Tethered flying flies change their visually-guided behavior after thermal conditioning

To explore whether and how flies adapt their behavior in response to aversive conditioning, we modified a well-established visual learning paradigm that was developed in the Heisenberg lab several decades ago [18]. In our modified paradigm, tethered flying flies in an LED arena were given closed-loop control of their angular orientation relative to a visual scene by locking angular rotations of visual patterns on the arena to differences in flies' left and right wingbeat amplitude, a proxy for their intended yaw movements (Fig 1a, upper panel) [37, 38] (see Methods). We used a periodic visual scene consisting of four quadrants of horizontal bars (Fig 1a, middle row). In two opposing quadrants, bars were positioned at a low elevation; in the other two quadrants, bars were positioned at a high elevation. We assessed flies' naive preferences for different quadrants of the visual scene during a pair of 2-min-long 'naive trials' (Fig 1a, bottom row). During subsequent 'training trials', two symmetric quadrants (the 'danger zone') of the visual scene were paired with an aversive heat punishment delivered via an infrared laser to the abdomen of the fly; the remaining two quadrants (the 'safe zone') were left unpunished. In 'probe trials' with no heat punishment, we assessed whether flies formed lasting associations between different quadrants of the arena and the aversive heat.

Prior to training, flies explored different parts of the visual scene (sample trajectories shown in Fig 1b, left column; Fig 1c, left column), but this changed during training trials with laser punishment, when flies typically avoided spending time in the danger zones (Fig 1b, middle column). In contrast, control flies that did not receive laser punishment continued to sample different parts of the scene ('no-laser' controls; Fig 1c, middle column). This trend continued in the probe trials, when laser-trained flies continued to avoid the danger zones even after the punishment was removed. Similar observations have been made in previous tethered fly visual learning studies in torque-meter-based flight arena setups [17, 18, 39] (see Methods for differences between these experiments and the setup we use here).

Past studies have shown that perturbing various CX neuron types significantly impacts the ability of flies to perform this operant learning behavior [27]. Furthermore, compass neurons [28, 31] and their inputs from the anterior visual pathway [40] are required for flies to display and maintain individualized heading preferences relative to a single visual landmark; this 'menotaxis' behavior is thought to aid dispersal and long-range navigation [41–46]. Additionally, inputs to the compass neurons have been linked to flies' ability to remember specific orientations relative to a disappearing visual landmark [36, 47]. We therefore sought to test whether flies' ability to flexibly

modify their naive preferences for different quadrants in our more complex visual setting also depended on the compass neurons.

Previous studies have quantified fly behavior in this operant visual learning task using a performance index ('PI score') [35, 48], which measures the relative fraction of time spent in safe versus dangerous quadrants; larger PI scores indicate a stronger preference for safety. We found, consistent with the sample trajectories (Fig 1b,c) and with past results [27, 35, 48], that laser-trained flies, on average, learn within minutes to avoid quadrants associated with punishment; this was reflected in an increase in PI scores during training trials that was maintained in probe trials (Fig 1e, left column, shaded bars). No-laser control flies did not significantly change their preference for either quadrant (Fig 1e, left column, white bars). We next tested flies in which we silenced compass neuron activity by selectively expressing the inwardly rectifying potassium channel, Kir2.1, in these neurons using two different split-GAL4 lines. Although these flies and flies from their parental control groups did not fly as well as wild type flies (data not shown), both control groups in these double-blind experiments (see Methods) trended towards avoiding the danger zone during training and showed good learning performance in at least the first probe trial (Fig 1e, middle and right columns, yellow/gray bars). In contrast, flies with silenced compass neurons failed to learn from laser training associated with the danger zone, showing consistently low PI scores (Fig 1e, middle and right columns, red-trimmed bars). These trends were reflected in the residency of these flies in safe versus danger zones in the last two probe trials relative to the first two 'naive' trials (Fig 1f). On average, wild type and parental control flies showed an increased residency in safe zones and reduced residency in danger zones after laser training. The averaged residency profiles of no-laser control flies and compass-neuron-silenced flies changed across these trials, but without the same increase specifically in the safe zone. These averaged behavioral trends were also visible at the level of individual flies. Individual wild type flies showed significantly higher PI scores after training (Fig 1g), which was not true of the no-laser control flies. Similarly, flies from one parental control group showed a significant increase in their PI scores while another trended in the same direction, but matching EPG-silenced flies did not show such increases in their PI scores. Thus, although the impaired flight performance of Kir flies limited our ability to assess their performance fully, we conclude that wild type flies quickly learn to avoid dangerous parts of the visual scene in our behavioral setting, and that compass neurons are required for normal visual learning in this operant behavior.

A probabilistic policy captures tethered flies' visually-guided behavior

What precise changes in behavior underlie this operant learning [17, 49]? To answer this, we sought to construct a generative model of behavior, or behavioral policy [16], that could account for naive and conditioned behavior. We noted that PI scores do not capture the different types of behavioral trajectories that flies generate in this paradigm, with or without conditioning (Fig 1b,c). Thus, our goal was to use a behavioral policy to characterize flies' behavioral variability in the absence of conditioning and to then predict how this variability should change based on experience.

Flies' behavior can be quantified in terms of the execution of different modes of patterned movement [50, 51]. During free and tethered flight, flies exhibit periods of fixation during which they maintain a near-constant heading over time [37, 50, 51]. These fixations are often punctuated by body 'saccades', or ballistic turns, that result in abrupt changes in heading [50, 51] (Fig 2a). We observed these behavioral modes in both laser-trained (Fig 1b) and no-laser control (Fig 1c) flies (note that there were no obvious differences in the distribution of fixation durations between laser-trained and no-laser control flies; see SI Fig S1). We approximated behavior as being composed of only these two modes (Fig 2a), and we used behavioral kinematics to determine transitions between them (SI Fig S2a; see Methods). Individual fixations tended to be long in duration with near-zero average angular velocity, while individual saccades were distinguished by high angular velocity over short durations (SI Fig S2b). We used the variability in these properties across flies and trials to infer a generative model of behavior in which flies control the distribution of possible actions within each of these two modes. Specifically, our analysis supports a generative model in which flies control the relative probability of initiating clockwise versus counterclockwise saccades through an adaptive rotational bias, and control the average duration of fixations through a drift-diffusion process with an adaptive drift rate (Fig 2b,c, SI Fig S2c-j; see Methods for a detailed analysis of additional control parameters).

We then asked how these two adaptive control parameters —rotational bias and drift rate— should be tuned as a function of the fly's orientation in the arena. Flies, like many other insects, display individual heading preferences, maintaining a specific 'goal heading' for periods of time even when tethered in visual virtual reality arenas (SI Fig S3) [28, 31, 40]. However, rather than purely fixating on one goal heading, both walking and flying flies also explore

other headings while centering their explorations around the goal heading [28, 31, 40, 52–54], a behavioral pattern that matches our observations in this paradigm. When model flies were trained through an RL algorithm to maintain a specific 'preferred arena heading' (see *SI: Reinforcement learning framework*), they minimized the drift rate of their fixations at that heading, and they learned to bias their saccades towards the corresponding orientation within the arena (Fig 2d, left column of e). This policy, which was optimized for a single preferred heading, produced qualitatively different trends in the average durations of fixation and directionality of saccades at $180°$-opposite headings (Fig 2e, right column).

We next sought to compare the behavioral patterns of real flies to those of the model flies. When we aligned data from individual laser-trained flies' fixations and saccades to their individual preferred headings in the arena (Methods), the behavior showed a bimodal rather than unimodal structure, with flies locally directing their saccades towards, and fixating longer at, the two headings that correspond to symmetric views of the visual scene (Fig 2f). Importantly, when we aligned behavioral data from flies that were not trained with laser punishment to their individual preferred headings in the arena, we found the same structure (Fig 2g). Similar observations of biomodal behavioral patterns have been made in a variety of different tethered flight arena studies of fly visual learning involving symmetric presentations of two visual patterns [17, 18, 39]. Flies' avoidance of the punished quadrants in probe trials in these experiments has been interpreted as evidence for their ability to associate punishment with specific visual patterns [55]. The symmetry of the visual scene and the potential confound that this presents to inferring heading in the scene might seem to rule out the possibility of flies using heading information for learning. However, our EPG silencing experiments implicate the compass system in this behavior (Fig 1). Additionally, flies' flexibility in heading preferences depends on compass neurons [28, 31, 40], and inputs to the compass neurons have been linked to flies' ability to remember specific orientations relative to visual patterns [36, 47]. Further, neurons upstream and downstream of compass neurons have been implicated in behavioral experiments similar to ours [27, 56]. Taken together, these results strongly suggest that flies rely on their heading representation, and likely on a behavioral policy tethered to this representation, to perform this visual learning behavior.

To understand how a compass-neuron- and heading-dependent policy might result in behavioral readouts that are bimodal in their dependence on arena heading, we considered the known instability of the fly heading representation in visual environments with two prominent features placed $180°$ apart [19, 29, 30]. That is, we wondered whether the observed bimodality in flies' behavior might arise from a similar instability in heading dynamics in our symmetric visual setting as well. Indeed, when we coupled the goal-heading-dependent policy to an unstable heading representation (Fig 2h; see Methods), the resulting behavioral readout exhibited a bimodal structure (Fig 2h) that mimicked the observed behavioral data (Fig 2f,g). The fact that we saw this structure in both laser-trained and no-laser control flies (Fig 2f,g) suggests that menotaxis-like behavior is governed by a 'pre-structured' internal policy tethered to the difference between flies' current internal heading and a single internal goal heading, and that all flies have a goal heading [28, 31, 40]. Taken together, these results motivated our conceptual model for the behavior (Fig 2i): the model assumes that a symmetric visual scene would be flexibly mapped onto the compass neuron population [29, 30] in a manner that creates a structured instability in heading dynamics, and that downstream circuits, which implement a pre-structured unimodal behavioral policy for saccades and fixations tethered to this HD representation, would inherit this instability. This, in turn, would produce the bimodality observed in behavioral profiles. By tethering a pre-structured policy to the difference between the fly's current and internal goal headings, the model ensures that any shift in the internal goal heading through learning would not change the bimodal structure of the profiles for fixation duration and turn bias (Fig 2i). Rather, these behavioral profiles would be shifted in heading space without losing their overall shape. When we examined the strength of flies' arena heading preference, this is indeed what we saw: both laser-trained and no-laser control flies showed a strengthening of their inferred goal heading (Fig 2j, top panel; see Methods). This strengthening may be the result of flies improving the precision and control of their saccades and fixations in this new visual setting; the structure of flies' behavior is influenced by the visual texture of their surroundings in free flight [57, 58]. Notably though, while both laser-trained and no-laser control flies increased the strength of their preferred headings across trials (see Methods), the laser-trained flies did so while also shifting their preferred headings significantly closer to the safe zone (Fig 2j, bottom panel). Thus, flies' observed behavior is consistent with a model in which a pre-structured, unimodal, heading-dependent policy controls flies' saccades and fixations.

A simple circuit model to implement a pre-structured goal-driven behavioral policy

How might our conceptual model for heading-dependent operant visual learning be implemented by the fly CX? To understand how the flies' circuitry could ensure that both the heading instability and the behavioral policy remain

structured with respect to the goal heading (Fig 3), and how learning acts to shift these structures (Fig 5), we constructed an abstract circuit model that expands upon the algorithm shown in Fig 2i. This model combines insights from existing models of the heading system [29, 59–64] with conceptual insights from the fly CX connectome [34], but it intentionally abstracts away much of the known detail of CX circuit structure and function to focus on key computations that we believe underlie flies' behavior in this assay. Below and in Fig 3, we break down the different components of the model and, in cases where CX neuron physiology and/or connectivity strongly suggest a role in such computations, link these components to potential CX neurons and circuit motifs (see SI Fig S4 and Supplemental Information for a more detailed description of specific CX neurons that might implement various aspects of the model).

The components we chose for our circuit model ensured sinusoidal distributions for the fly's behavioral output as a function of angular distance from its internal goal heading, matching the behavior of both laser-trained and no-laser control flies (Fig 2f,g) and also matching the RL-based optimization of model fly behavior in our task setting (*SI: Reinforcement learning framework*). Our model assumed that neurons in the PB shape the compass neuron's HD signal into a sinusoidal shape [34, 65], a profile ideally suited for performing vector computations [66]. We did not impose this constraint on the goal activity vector, which we assumed is built into an arbitrary shape through experience in a downstream CX structure, the fan-shaped body (FB) (see below). The model assumes that the fly's motor drive is derived from the overlap between this arbitrarily-shaped goal activity vector and the sinusoidal compass vector (Fig 3a). Mathematically, the summation of a point-wise multiplicative operation between the compass and goal vectors ensures that the resulting motor drive profile is itself sinusoidal across the range of heading differences, regardless of the specific profile of goal activity (Fig 3b); thus, the circular mean of the goal activity specifies the goal heading (see Eq. 31 in *SI: Reinforcement learning framework* for a brief mathematical explanation). With such an operation, a weaker goal vector produces a reduced motor drive that nevertheless maintains this sinusoidal profile over the range of differences between current and goal heading (Fig 3b, rightmost column). Overall, this scheme ensures that the shape of the motor drive is sinusoidal regardless of the absolute orientations of the two vectors. How might such a sinusoidal motor drive be converted into an appropriately phase-shifted sinusoidal motor drive to control the drift rate of fixation and the turn bias in saccades, the parameters that appear to control the actions of tethered flying flies? To produce the lowest drift rate at the fly's internal goal heading, the motor drive curve would need to be shifted by $180°$; to produce the strongest turn bias towards the internal goal heading when the fly is oriented orthogonal to its goal heading, the motor drive curve would have to be shifted by $90°$. Our circuit model assumed that the heading population vector is itself shifted appropriately before it is converted into motor drive based on its overlap with an internal goal vector (Fig 3c). Shifts by $90°$ in either direction control saccades, producing the largest output —and thus most likely driving CW or CCW saccades— when the fly's current heading bump is $90°$ to the right or the left of the goal heading, respectively. A shift by $180°$ has the largest output —and thus likely drives short fixations with a high drift rate— when the fly's current and goal headings are anti-aligned. Together, this architecture ensures that the fly's behavior remains structured with respect to the angular difference between the current and goal headings, regardless of their absolute orientations. Indeed, when we simulated the circuit model using these phase shifts and multiplicative operations for action and goal activity, it produced behavioral trajectories and distributions of fixation duration and turn bias that were qualitatively similar to those we observed in real flies (compare Figs 1b,c and 2a to the upper panel of Fig 3d). The shape and range of the goal activity determines how strongly the goal heading drives fixational and saccadic behavior; the stronger the circular mean of the goal activity profile, the larger the difference in behavior at the goal and anti-goal headings, and thus the more structured the behavioral output (compare Fig 2f,g to the lower panels of Fig 3d).

Putative CX output neuron types that seem ideally suited to carrying out exactly such phase shifts of the heading population vector have been described by us and others [34, 62, 67]. Phase shifts in the CX are determined based on propagating a potential activity bump carried by the compass neurons to their downstream partners in the PB and the FB ([34, 65]). Neurons that receive input in the PB from similarly-tuned compass neurons and project to similar columnar regions of the FB are considered to have a $0°$ phase shift (these phase shifts motivate the coloring of PB glomeruli and FB columns in Fig 3e) (see [34] for details). Based on such phase propagation and also on their anatomical projection patterns, the neurons most likely to serve the function of phase-shifted action neurons for saccades are the left and right PFL3 neuron populations. Each population inherits the heading bump from the compass and other neurons in the PB. However, we assume that these neurons participate in goal-related computations in the FB, where they project with $-90°$ and $+90°$ phase shifts relative to their PB inputs (Fig 3e, left and right columns) (see [34] for details of phase assignments). Further, both the left and right PFL3 neuron types project to a premotor region called the LAL, where they are well-positioned to drive downstream controllers for CW or CCW rotations (see LAL projections schematized in Fig 3e, left and middle columns) [32, 34, 62, 67]. A third

population of PFL2 neurons receives a version of the current heading bump that is phase-shifted by 180° [34] in the FB and projects bilaterally to both controllers, thereby potentially controlling fixation ([34]; see also [67]) (Fig 3e, left and middle columns). The PFL2 and PFL3 populations mutually inhibit each other, to prevent saccades during a fixation, and vice versa. In summary, our anatomically-inspired model assumes that heading and goal activity is combined in PFL2 and PFL3 neurons whose net output is used to first determine whether to fixate or saccade, and, next, to control the duration of fixations or the direction of saccades as appropriate.

Anatomical data from the CX provided clear candidates for action neurons that might integrate heading and goal information to mediate fixations and saccades. By contrast, a far greater number of neuron types with columnar projections in the FB could each represent a potential goal vector that is ultimately combined with heading activity in the three PFL neuron populations [34]. Our model assumes that goal weights are stored in excitatory synapses between motor-state-activated FB 'tangential neurons' ([34]), which project to all FB columns within a layer, and as-yet-unidentified FB columnar 'goal neurons' that have a 0° phase shift and that, in turn, synapse onto PFL2 and PFL3 neurons (Fig 3f, bottom left; see *SI: Linking the Conceptual Model to Known Anatomy* for potential neuron types that could serve such functions). During flight, the goal neurons read out the strength of the goal weights and, through their population activity, represent the goal vector that is then multiplied by the phase-shifted heading vector in PFL2 and PFL3 neurons to determine the fly's actions. During operant learning, the fly receives laser punishment in the danger zone, which, in the model, is carried to the FB by the activity of a different class of neuromodulatory and potentially dopaminergic FB tangential neurons (Fig 3f, bottom right; see *SI: Linking the Conceptual Model to Known Anatomy* for candidates). We hypothesize that these neuromodulatory neurons receive local heading input near their output synapses from 0° phase-shifted FB columnar neurons, an idea with anatomical support from the connectome [34]. Thus, the punishment signal carried by these neurons would be transformed into a heading-shaped bump that triggers bidirectional Hebbian plasticity at synapses between co-active motor-state and goal neurons (Fig 3f, top). Thus, negative reinforcement moves the goal heading away from the current compass heading, and positive reinforcement moves it towards the compass heading (Fig 3g). However, the fixation duration and turn bias curves maintain their structure even as the goal heading evolves (Fig 3h). As the goal heading strengthens, the fly's behavioral expression of its internal goal also grows stronger, with longer fixation durations near the goal and stronger turn biases towards it (Fig 3h, rightmost columns). Within the constraints of the learning algorithm considered here, a high degree of behavioral structure is achieved by driving the goal weights towards a sinusoidal profile (see Eq. 31 in *SI: Reinforcement learning framework* and the following discussion for a brief mathematical explanation). In summary, we propose that tangential and columnar neurons in the FB store and—in the presence of heading-shaped reinforcement from neuromodulatory FB tangential neurons—update goal weights in their synapses. These weights are read out by other columnar neurons before being transmitted to action neurons, which combine this input with phase-shifted heading information in a manner that enables flies to maintain structured behaviors around internal goals even as they change. When the learning rule we use is coupled with the structured policy implemented by this circuit model, the model fly's goal heading strengthens and moves towards safety (Fig 3i), matching our behavioral results (Fig 2j).

Visual symmetries trigger jumps in flies' internal HD representation

Our heading-tethered policy produces bimodal curves for fixation duration and saccade bias only in the presence of specific instabilities in the heading representation (Fig 2h). Thus, a key issue for our circuit model to address is how the symmetric visual setting of our task influences the computation of heading itself. The fly's current HD is maintained as a single bump of activity that typically moves in concert with the fly's rotations [19]. We assumed that the bump maintains a von Mises profile in the EB before it is transferred to other neurons as a sinusoid [34, 65], and that its responsiveness to the fly's saccades are accurately maintained by a ring attractor network and by neuron types that we did not explicitly model [29, 34, 40, 61, 65, 68, 69]. We focused instead on the stability of this heading representation, which, in our model, is determined by a set of plastic synapses from visual ring neurons to compass neurons [29, 30, 63, 70] (Fig 4a). Visual ring neurons have feature-tuned receptive fields that tile space (Fig 4a, upper panel; [71–73]), and they synapse onto compass neurons via all-to-all inhibitory connections in the EB [34] (Fig 4a, lower left panel; also see *SI: Linking the Conceptual Model to Known Anatomy*). During exploration of a visual scene, inhibitory Hebbian-like plasticity is thought to weaken synapses from active ring neurons onto active compass neurons at the location of the heading bump in the EB. We assume that the plasticity rule depends on angular velocity [29], a dependence that we model as acting through hypothesized angular-velocity-dependent neuromodulatory neurons (Fig 4a, middle row circuit schematic) that innervate the EB [34, 74, 75] (see *SI: Linking the Conceptual Model to Known Anatomy* for candidate dopaminergic neurons that project from motor centers

to the EB). Over time, this plasticity creates a consistent mapping between the visual scene and the heading representation, which in a simple scene with a single landmark would be characterized by a diagonal band in the matrix of synaptic weights between ring and compass neurons (see heatmaps in Fig 4b; [29, 30]). When stabilized, this mapping would enable the bump to move around the EB in perfect synchrony with the rotation of the visual scene, which would be captured by the activation of ring neurons with the appropriate spatial receptive fields (Fig 4b).

We next asked what sort of mapping the same plasticity rule would produce in a symmetric scene. In a previous study, we showed that the stability of the mapping between a visual scene and the heading representation depends on particular characteristics of that scene; briefly, scenes whose rotational auto-correlation produce a single, dominant peak tend to induce stable, one-to-one mappings from the visual scene onto the heading representation [29]. In contrast, the visual scene employed in our behavioral experiments is two-fold symmetric, with two peaks in its rotational auto-correlation. We next simulated the mapping of a simpler scene with such twofold symmetry, one in which there were two vertical features, $180°$ apart. As the fly moves through the scene, the mapping between ring and compass neurons develops multiple bands so that ring neurons with $180°$-opposite receptive fields have approximately equal synaptic weights onto the same compass neurons (Fig 4c). Importantly, when this mapping is driven by behavior that is tethered to an internal goal heading, the peaks and troughs (bright and dark patches) in this mapping develop at specific locations relative to the goal (Fig 4d, top right). Specifically, because most of the fly's saccades are in the neighborhood of the goal heading, the compass headings around the internal goal heading are also the most stable; that is, compass neurons tuned to headings near the goal heading experience the weakest inhibition from the $180°$-symmetric ring neurons activated by the visual scene at those headings (Fig 4c; d, top right), and compass neurons with $180°$-symmetric tuning receive the strongest inhibition from the same ring neurons. Since the scene is symmetric, these ring neurons will be identically active at two $180°$-symmetric arena headings, but will inhibit compass neurons with $180°$-opposite HD tuning to different degrees. The panels in Fig 4c play out the scenario that results as a fly with such a mapping saccades through the symmetric scene. The fly starts with its compass bump in a position of weakest inhibition (the most stable position for the bump), but as it saccades to the opposite side of the visual scene (that is, to a view identical to its initial starting point), the compass bump moves to a position in which it is strongly inhibited by active ring neurons. This triggers an instability in the dynamics of the network [19, 29, 30, 76], and the bump jumps across the EB to the compass neurons that are most weakly inhibited by the activated ring neurons (Fig 4c, right column; bump jumps from brighter to darker parts of the EB).

As discussed in the previous section, synaptic weights from individual ring neurons onto compass neurons influence bump instability at different orientations within the EB (Fig 4c). More precisely, it is the summed (net) inhibition from the population of active ring neurons that determines the compass bump's relative stability at any EB orientation: the bump will preferentially occupy regions of the EB that are only weakly, rather than strongly, inhibited by active ring neurons (Fig 4c). Note that such instabilities do not necessarily require symmetric scenes; the presence of multiple similar visual features in a given scene would evoke similar ring neuron activation patterns at multiple headings, and this similarity in ring neuron activation patterns would then trigger a competition in multiple compass neurons that are tuned to those headings. In the specific case of a two-fold symmetric visual scene, our circuit model assumes that the difference in net inhibition between two locations in the EB separated by $180°$ determines the probability that the bump will jump between these internal compass headings (Fig 4d, red curves to the right of weight matrices). If the ring-neuron-to-compass-neuron synaptic weights are stabilized and strong, as would result from a strong goal heading that drives the bump's movements to the neighborhood of the goal heading in the EB, the difference in inhibition between EB locations $180°$ away from each other would be greater, leading to an increased probability of the bump jumping away from an EB location with strong inhibition to the one that is more weakly inhibited. Over time, the heading instability will produce a sinusoidal jump probability curve that is tethered to the policy, such that the most stable bump heading —the EB location where the bump is least likely to jump— is aligned with the goal heading (Fig 4d, upper right; note how the red bump jump curve is aligned to the blue goal heading). This, in turn, results in a bimodal behavioral readout (Fig 4d, lower right). A stronger goal heading thus produces stronger mappings of the visual scene onto the compass system, an increased probability of jumps in symmetric environments, and, as described above, stronger behavioral readouts.

We next ran our circuit model on this simple symmetric scene with two vertical stripes placed $180°$ away from each other and with a single internal goal heading (Fig 4e, left panels). After the mapping from ring to compass neurons stabilized, the compass bump moved in concert with the fly's saccades and remained tethered to changes in the orientation of the visual scene, but with occasional $180°$ jumps towards the goal heading (Fig 4e, middle panel). This was captured by the fraction of time that the bump spent with different angular relationships to the visual scene, a

parameter that we and others refer to as the 'visual offset' [19, 29–31, 76] (schematized in Fig 4e, left bottom panel). This offset is arbitrary, varies from fly to fly, and can vary over time as well [19, 28–31]. We found that over the course of the simulated trial, the model circuit's HD bump was at one of two different visual offsets that were separated by $180°$, which would correspond to identical views of the visual scene (Fig 4e, right panel). We predicted that bump dynamics in the two-fold symmetric scene used in our behavioral paradigm and in many visual learning experiments before it [18, 27] would resemble the unstable dynamics seen in our simulations, even though the scenes differ in the number of visual patterns (two vertical stripes in these simulations and four horizontal stripes at two different heights in the paradigm that real flies experienced).

To ask if the real fly's heading dynamics exhibit a structured instability in the symmetric visual setting used in our behavioral paradigm, we monitored the compass neuron HD representation in the EB (Fig 4a, middle row) using two-photon calcium imaging in tethered flies flying in a visual setting similar to that used in the learning assay (Fig 4f, left column; see Methods for details). When we analyzed the stability of the offset of the heading representation relative to the visual scene, we found, as predicted, that the heading bump tended to jump between two offsets that reflected symmetric views of the scene (Fig 4f, middle column). In our two-fold symmetric scene, this corresponds to a jump of $180°$ (Fig 4f, panels at right). Examining offsets across flies (Fig 4g), we found that the distribution of offsets was bimodal in a majority of flies (Fig 4g, bottom left panel), with peaks separated by $180°$ (Fig 4g, bottom right panel; see *SI: Linking the Conceptual Model to Known Anatomy* for discussion of the smaller peak at $90°$). In correspondence with this, we found that different wedges of the EB were active at symmetric angular orientations, and thus their heading tuning curves had two peaks also separated by $180°$ (Fig 4h). This resulted in a two-to-one mapping from the visual scene onto the heading representation, similar to the tuning previously observed in simpler symmetric scenes [19, 29, 30, 76]. We next asked whether bump jumps were more likely to occur at certain locations in the EB (Fig 4i). We measured how frequently the bump tended to jump from different locations in the EB, relative to the number of visits the bump made to that location. When we analyzed the location of these jumps relative to the location of maximal residency in the EB, we found that jumps were least likely to occur at this inferred goal heading, and most likely to occur $180°$ away from this inferred goal heading (Fig 4i, bottom panels). Together, this suggests that the symmetries of the scene induce instabilities [19, 29, 30, 76] that manifest in jumps of the heading bump between locations in the EB that correspond to symmetric views of the visual scene. In a scene with two-fold symmetry, this instability is not uniform around the EB, but is strongest at the location symmetric to the preferred bump heading (Fig 4i, bottom panels), as predicted (Fig 4c,d). Taken together, these results are consistent with the observed structure in flies' behavior (Fig 2f,g) resulting from the impact of the scene's symmetry on the dynamics and stability of the heading representation, and of a behavioral policy tethered to that representation.

Deciphering fly-to-fly variability in operant visual learning

In our CX-based model of how the fly behaves in our visual learning paradigm, the fly must form a new goal heading alongside a new compass representation to which the evolving goal heading is tethered. The mapping of the visual scene onto compass neurons is an unsupervised learning process modulated by the fly's saccades through the visual scene. On the other hand, the formation of a goal heading is a supervised learning process guided by the positive and negative experiences that the fly has at different internal headings. How is a fly's behavior influenced by the interactions between the two simultaneously occurring learning processes, and how do the two internal representations —of compass heading and goal heading— influence each other and ultimately the fly's actions? To answer these questions, we ran a series of simulations using our circuit model. For each model fly going through the simulated task, we kept track of the most stable compass heading (Methods) and the internal goal heading, both relative to the arena's safe and danger zones (Fig 5a); we also tracked the strength of the fly's evolving internal goal. To ensure that each model fly had similar starting conditions, we began each simulation by allowing flies to create a stable, single-banded compass mapping and a goal heading in the middle of what would later become the danger zone (Fig 5b, first column; see Methods). At this point, the probability of the bump jumping is nearly uniform across the EB. We then changed the scene to be two-fold symmetric as in our behavioral paradigm (see Methods), with negative and positive reinforcement in danger and safe zones, respectively. The structure of flies' exploration of the arena and of different bump headings is at first dictated by their initial compass mapping and goal heading (Fig 5b, first column). Over time, the compass mapping changes to become double-banded, and the jump probability curves and goal activity vectors become stronger, more sinusoidal, and more correlated with one another, eventually locking to one of the two safe zones (see evolution of mapping and profiles across the columns from left to right in Fig 5b). Different flies go through different experiences because of the probabilistic nature of the structured behavioral policy. However, on average, model flies increase the strength of their goal headings and shift

their goal heading towards safety, which results in an increase in their PI scores (Fig 5c). We next compared the performance of flies that began with different strengths of goal heading. We found that flies took different trajectories through the space of goal headings before they evolved strong goal headings centered on the safe zone (Fig 5d, top panel). Even though the flies all began with goal headings centered on the danger zone, a stronger goal heading seemed to ensure a faster convergence to the final goal heading in the safe zone (Fig 5d, top panel). When we compared flies' most stable compass headings with their goal headings, we found that a good alignment between the two headings also seemed to allow flies to more rapidly converge onto strong goal headings in the safe zone (Fig 5d, second row of panels). In contrast, flies that had misaligned goal and stable compass headings seemed to get 'stuck' in bad parameter regimes and required more time to 'escape' to regimes that allowed them to discover safe goal headings (Fig 5d, bottom row of panels). If flies began with strong goal headings, they tended to maintain the strength of their goal heading throughout (Fig 5e). When goal headings were weaker (brightest regions in Fig 5e heatmap), they were more likely to suffer from misalignments between their most stable compass heading and their goal heading (Fig 5f), which, in turn, can adversely impact their performance in the task (see evolution of PI scores in Fig 5g). Thus, the fact that early learning in our paradigm requires a combination of mapping a novel visual scene onto the compass while simultaneously using this compass to learn a goal and navigate towards it with a structured but probabilistic policy creates dependencies between these different learning processes and representations. Flies' individual actions determine how the learning process evolves, but starting with a strong goal heading, even if it is inaccurate for the new environment, may structure behavior in such a way that it enables a faster remapping of the visual environment onto the compass, and a faster determination of good headings within the environment.

DISCUSSION

We analyzed *Drosophila* behavior in a modified version of a well-established visual learning paradigm [17, 18], and combined this analysis with calcium imaging of compass neurons in a similar visual setting to propose an anatomically-inspired circuit model of CX computations underlying this operant behavior. We did not aim to include all the available details of CX circuitry, instead incorporating only those aspects of CX anatomy and physiology that we believe to be essential to understand circuit function in our task. This abstract model enabled us to examine how individual flies use evolving heading representations to guide their behavior towards goals that they simultaneously infer from heading-specific heat reinforcement. Based on our behavioral and physiological results, as well as previous physiological, behavioral, connectomic, and modeling studies focused on the CX [19, 28–32, 34, 59, 60, 62, 63, 67, 68, 76], we suggest that flies rely on a pre-structured behavioral policy tethered to a flexible internal heading representation that specifies controllable properties of the fly's actions relative to a single goal heading. Flies operantly learn to shift this goal heading based on reinforcement, quickly redirecting their actions toward unpunished headings. However, because the heading representation on which this policy is built is learned alongside the goal heading, any instability in this representation impacts the fly's behavior, which, in turn, impacts the fly's ability to sample the new environment and appropriately update its goal heading. We used artificial visual environments with repeating patterns tethered to aversive punishment to reveal this interdependence of representational stability, goal learning, and goal-directed behavior. Our framework makes clear how symmetries in the visual scene induce the fly's heading bump to jump, and how this instability, when appropriately structured, does not interfere with the fly's ability to learn in this visual setting. Because we study the learning process beginning with the fly's first encounter with this sensory environment, the formation of the heading representation is intrinsically coupled with the formation and shifting of the goal heading. As a result, instabilities in the heading representation can impact learning, something that we see play out at the level of model flies' changes in performance.

Several recent models have relied on PFL3 phase shifts, or asymmetries in synapse counts from EPG to PFL neurons in the PB, to move model flies to goal headings [32, 62, 64, 67]. In these models, flies direct turns of appropriate size to reach goal headings, with any deviations from such turns considered noise. In our formulation, the fly's brain builds in a trade-off between exploration and exploitation through an intrinsically probabilistic policy, and CX neurons control parameters of this distribution. We find that this approach better captures the finer-grained aspects of flies' behavior in our experiments, specifically, the dependence of saccade bias and duration of fixation on heading. The potential mechanisms we propose for how potential goal neurons might operate, and for how 'goal weights' might be stored, updated, and read out, do not yet have strong support in available physiological and anatomical data, but there is evidence that an FB columnar neuron type called hΔc, which carries an odor-gated wind-direction signal, may serve as a goal heading during olfactory navigation [67]. Additional experiments will be needed to ascertain whether these or other FB columnar neurons have the response properties required for them to perform the computations that we hypothesize in our model.

Our results may warrant a reinterpretation of decades of studies in visual pattern, spatial, and color learning in tethered flying flies [18, 27, 35, 49]. In all these studies, flies were believed to have learned to associate their actions with specific visual features or objects; these conclusions were based on the near-symmetric structure of fly behavior in settings with the same patterns in opposing quadrants of circular arenas. This symmetry of the visual environment was believed to rule out the possibility of learning based on heading. We suggest instead that flies build heading representations tethered to these different visual surroundings, and rely on heading-representation-based goal learning to associate rewards or punishments with different headings. We further argue that the observed bimodality in flies' responses to visual patterns were caused by an instability in the heading representation that arose from a symmetry of the scene. This structured instability rendered irrelevant any limitations that a single-goal-heading-based policy might otherwise impose on the fly's actions in a setting with multiple safe and dangerous headings. It is possible that such heading instabilities would be less frequent in free flight, where proprioceptive cues are likely to play a greater role in controlling heading bump dynamics (see, for example, [76]). Notably, neurons encoding the symmetries of an animal's environment are not limited to tethered insects: neurons in the retrosplenial cortex of freely behaving rodents also display such responses [77], and bidirectional HD tuning in that system has also been modeled as relying on Hebbian plasticity [78]. Although the symmetric environments used in our studies and in related studies in other species may seem artificial, natural scenes can have repeating visual features, and these can trigger instabilities in HD dynamics as well [29]. The specific impact of such instabilities on the performance of particular tasks depends on how reliant downstream circuits are on heading-dependent policies. In the case of flies, it is not known whether or not they can learn multiple distinct goal headings, nor to what extent they can form more complex policies beyond the pre-structured single-goal-heading policy invoked here. Performance in a place-learning paradigm suggests that flies can learn more complex associations to guide navigation through 2D environments [79]. It is also possible that, in contrast to the CX-based learning we study here, many spatial navigation behaviors may rely on associations made in the mushroom body [23, 80–86], a brain region that has also been suggested to be involved in some operant visual learning tasks in tethered flying flies [35, 87].

For most of our analyses, we decomposed the tethered fly's behavior into two different modes: fixations and saccades. Freely flying flies are known to exhibit these different modes [88] that are characterized by distinct kinematic properties and necessitate both continuous and discrete control [51, 89]. The same modes have also been observed previously in tethered flying flies [51]. We explicitly incorporated these modes into the construction of a behavioral policy, and we used the observed variability in kinematic parameters across flies and trials to infer the parametric form and control parameters of this policy. When combined with optimal RL algorithms [16], this enabled us to specify how these control parameters should change based on experience. This approach bears similarities to recent studies proposing that learning operates on generative parameters that control distributions of movements, rather than on the higher-dimensional space of all possible movements [90–92]. We then used this approach to specify how these control parameters should be structured as a function of the fly's current heading to maintain a goal heading over time. Rather than learning this structured relationship from scratch, we showed that this relationship is pre-built into how untrained flies sample their surroundings, a strategy that might facilitate dispersal in the absence of explicit goals [52]. Indeed, these same patterns of structured behavior resemble those observed in tethered behaving flies responding to visual features that are innately attractive or aversive [89]; note that an additional parameter, the size of saccades, also varies based on angular distance from a 'goal' object in those visual environments, something that was less striking our visual setting (data not shown). We suggest that flies may, in fact, rely on different visuomotor pathways in different settings. Responses to innately attractive or aversive objects [93–97] could rely on direct and hardwired visuomotor pathways that recruit banks of feature detectors in the optic foci [98, 99] and, perhaps, relatively stereotyped motor responses dependent on the spatial receptive fields of feature detector inputs [89]. In contrast, our compass-neuron silencing experiments suggest that learned responses rely on an indirect pathway that recruits pre-structured probabilistic behavioral policies tethered to the relationship between internal representations of current and goal headings. We did not explore whether these probabilistic biases could be strengthened through longer training protocols. We note that although it might seem optimal to steer towards and then maintain a single goal heading rather than initiate directed saccades that are probabilistically biased toward this heading, using such a default behavioral strategy would likely be too predictable to avoid predation [100] and would minimize exploration. Indeed, many animals, including flies, show stochasticity in their actions when behaving freely in dynamic settings [100, 101]. In the fly, recent evidence from the connectome suggests that the architecture of columnar neurons in the FB could implement a pre-structured behavioral policy for steering towards a goal heading by using the fly's current bump heading [34, 62] (see also *SI: Linking the Conceptual Model to Known Anatomy*). Importantly, within this model, learning acts to modify the location and strength of the goal heading while preserving the entire structured relationship between different control parameters across different headings. Our results suggest that these associations are mediated via a flexible pathway through

the CX; however, direct sensorimotor pathways that instruct reflexive actions might, in fact, work in concert with these flexible pathways, and in the aversive conditioning setting considered here, might enable the fly to quickly escape punished zones. How such pathways are balanced to guide reflexive and flexible actions, and whether the outcome of reflexive actions can be used to inform future flexible actions, is not yet known (but see *SI: Linking the Conceptual Model to Known Anatomy* and *SI: Reinforcement learning framework* for a discussion of how reflexive actions might be incorporated in the framework presented here).

In contrast to many behavioral paradigms in mammals, flies in this paradigm learn within a matter of minutes —without shaping or instruction— to direct their behavior away from punishment and towards safety. Our results suggest that flies' rapid learning relies on a strong inductive bias in the form of a pre-structured behavioral policy that dictates flies' sampling strategy. This pre-structured policy effectively assumes the existence of a single goal heading for the fly, and efficiently directs them towards identifying and orienting towards such a heading, wherever that happens to be. Such inductive biases reduce the possible scenarios that are explored during learning and can thereby speed up the learning process when these scenarios are compatible with the learning task [102, 103]. Recent RL studies have explored how such inductive biases might be constructed by learning common features across different learning tasks [104, 105], a process known as learning to learn [106]. Here, we show how an inductive bias that is likely learned over evolutionary timescales can be inferred directly from an animal's behavior in the absence of an explicit task. The ability to rapidly exploit this inductive bias, in this case by shifting a single goal heading, relies on faster-timescale learning. Whether, and to what extent, this faster-timescale learning could modify the flies' pre-structured behavioral policy —for example, by suppressing the exploratory component through increased training— remains unknown, as does the potential for behavioral state information —for example, walking instead of flight— to switch which actions are controlled through the same behavioral policy.

It has recently been suggested that rapid learning in both artificial and biological systems relies on combining context-dependent memories with efficient exploitation of environmental and task structure [103]. Indeed, structural priors may hold the key to animal learning [107]. Here, we provide insights into how specific neural circuits might instantiate a behavioral policy that has evolved to address ecological needs through efficient actions, and how this policy both informs and is shaped by a flexible, and perhaps context-dependent, internal representation of the environment. Targeted genetic access to the specific cell types that might mediate this learning provides an avenue for rigorously testing these ideas in the future.

ACKNOWLEDGEMENTS

We thank Jason Wittenbach and Daniel Barabasi for pilot modeling efforts, Parvez Ahammad for useful early discussions, and Alice Wang for pilot experiments. Eyal Gruntman, Michael Reiser, Josh Dudman, John Tuthill, TJ Florence, Sung Soo Kim, Yoshi Aso, and members of the Jayaraman and Hermundstad labs provided insightful input at different points of the study. We received useful feedback on the manuscript from Sandro Romani, Marcella Noorman, Hannah Haberkern, and Dan Turner-Evans. We are grateful to Bjorn Brembs for informative email discussions on the design and interpretation of well-established visual learning studies, and to Stephen Thornquist and Gaby Maimon for helpful discussions on the statistical analysis of behavioral data in an earlier version of the manuscript. We thank Gudrun Ihrke for supporting these experiments through her expert management of Project Technical Resources (PTR). We thank Dan Milkie (now at Janelia) and Andy Chiu from Coleman Technologies for help with developing the FPGA Wingbeat Analyzer. We thank Janelia Experimental Technology (jET) for technical assistance, especially: Jinyang Liu for the LED arena controller code, Steve Sawtelle for the D2A converter connected to the FPGA Wingbeat Analyzer, and Tanya Tabachnik, Igor Negrashov, and Bill Biddle for designing and manufacturing the fly mounting assembly used for two-photon imaging. We are grateful to Janelia's Drosophila Resources team for stock building and maintenance, and to the Media Prep Facility for special fly food that kept our finicky flies flourishing.

This work was funded by the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

CD and VJ conceived of the study waay back when. AMH, CD, and VJ then reconceptualized the study. RK performed all silencing experiments and associated control experiments under CD's supervision. CD performed all other experiments, data processing and initial data analysis. AMH performed all behavioral analysis in this

manuscript, with input from CD, VJ, and BKH. VJ and AMH analyzed imaging data with input from CD. All authors interpreted results. AMH developed the theoretical framework and performed all modeling and simulations, with conceptual input from VJ, CD, and BKH. The proposed CX circuit implementation of the model was conceived over multiple feedback loops involving VJ, AMH, CD, and BKH, with BKH contributing, in particular, to the CX implementation of the behavioral policy. AMH and VJ wrote the paper, with input and editing from CD and BKH.

COMPETING INTERESTS

We do not have any competing interests.

References

1. Dupre, C. & Yuste, R. Non-overlapping Neural Networks in Hydra vulgaris. *Current Biology* **27,** 1085–1097 (2017).
2. Wadhams, G. H. & Armitage, J. P. Making sense of it all: bacterial chemotaxis. *Nature Reviews Molecular Cell Biology* **5,** 1024–1037 (2004).
3. Huston, S. J. & Jayaraman, V. Studying sensorimotor integration in insects. *Current Opinion in Neurobiology* **21.** Sensory and motor systems, 527–534 (2011).
4. Calhoun, A. J. & Murthy, M. Quantifying behavior to solve sensorimotor transformations: advances from worms and flies. *Current Opinion in Neurobiology* **46.** Computational Neuroscience, 90–98 (2017).
5. Crochet, S., Lee, S.-H. & Petersen, C. C. Neural Circuits for Goal-Directed Sensorimotor Transformations. *Trends in Neurosciences* **42,** 66–77 (2019).
6. Pouget, A & Snyder, L. Computational approaches to sensorimotor transformations. *Nat Neurosci* **3 Suppl,** 1192–1198 (2000).
7. Wolpert, D. M. & Flanagan, J. R. Computations underlying sensorimotor learning. *Current Opinion in Neurobiology* **37.** Neurobiology of cognitive behavior, 7–11 (2016).
8. Knierim, J. J. & Zhang, K. Attractor dynamics of spatially correlated neural activity in the limbic system. *Annual Review of Neuroscience* **35.** PMID: 22462545, 267–285 (2012).
9. Finkelstein, A., Las, L. & Ulanovsky, N. 3-D Maps and Compasses in the Brain. *Annual Review of Neuroscience* **39.** PMID: 27442069, 171–196 (2016).
10. Tolman, E. C. & Honzik, C. H. Introduction and removal of reward, and maze performance in rats. *University of California Publications in Psychology* **4,** 257–275 (1930).
11. Huber, D. *et al.* Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* **484,** 473–478 (2012).
12. Poort, J. *et al.* Learning Enhances Sensory and Multiple Non-sensory Representations in Primary Visual Cortex. *Neuron* **86,** 1478–1490 (2015).
13. Peters, A. J., Lee, J., Hedrick, N. G., O'Neil, K. & Komiyama, T. Reorganization of corticospinal output during motor learning. *Nat Neurosci* **20,** 1133–1141 (2017).
14. Coddington, L. T. & Dudman, J. T. Learning from Action: Reconsidering Movement Signaling in Midbrain Dopamine Neuron Activity. *Neuron* **104** (2019).
15. Kuchibhotla, K. V. *et al.* Dissociating task acquisition from expression during learning reveals latent knowledge. *Nat Commun* **10,** 2151 (2019).
16. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* Second (The MIT Press, 2018).
17. Brembs, B. & Heisenberg, M. The operant and the classical in conditioned orientation of Drosophila melanogaster at the flight simulator. *Learn Mem* **7,** 104–115 (2000).
18. Wolf, R. & Heisenberg, M. Basic organization of operant behavior as revealed in Drosophila flight orientation. *Journal of Comparative Physiology A* **169,** 699–705 (1991).
19. Seelig, J. D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521,** 186–191 (2015).
20. Strauss, R. The central complex and the genetic dissection of locomotor behaviour. *Current Opinion in Neurobiology* **12,** 633–638 (2002).
21. Pfeiffer, K. & Homberg, U. Organization and functional roles of the central complex in the insect brain. *Annu Rev Entomol* **59,** 165–184 (2014).
22. Turner-Evans, D. B. & Jayaraman, V. The insect central complex. *Curr Biol* **26,** R453–457 (2016).
23. Webb, B. & Wystrach, A. Neural mechanisms of insect navigation. *Curr Opin Insect Sci* **15,** 27–39 (2016).
24. Varga, A. G., Kathman, N. D., Martin, J. P., Guo, P. & Ritzmann, R. E. Spatial Navigation and the Central Complex: Sensory Acquisition, Orientation, and Motor Control. *Front Behav Neurosci* **11,** 4 (2017).
25. Heinze, S. Unraveling the neural basis of insect navigation. *Curr Opin Insect Sci* **24,** 58–67 (2017).
26. Honkanen, A., Adden, A., da Silva Freitas, J. & Heinze, S. The insect central complex and the neural basis of navigational strategies. *J Exp Biol* **222** (2019).

27. Liu, G. *et al.* Distinct memory traces for two visual features in the Drosophila brain. *Nature* **439,** 551–556 (2006).
28. Giraldo, Y. *et al.* Sun Navigation Requires Compass Neurons in Drosophila. *Curr Biol* **28,** 2845–2852 (17 2018).
29. Kim, S. S., Hermundstad, A. M., Romani, S., Abbott, L. F. & Jayaraman, V. Generation of stable heading representations in diverse visual scenes. *Nature* **576,** 126–131 (2019).
30. Fisher, Y. E., Lu, J., D'Alessandro, I. & Wilson, R. I. Sensorimotor experience remaps visual input to a heading-direction network. *Nature* **576,** 121–125 (2019).
31. Green, J., Vijayan, V., Mussells Pires, P., Adachi, A. & Maimon, G. A neural heading estimate is compared with an internal goal to guide oriented navigation. *Nat Neurosci* **22,** 1460–1468 (2019).
32. Rayshubskiy, A. *et al.* Neural circuit mechanisms for steering control in walking Drosophila. *bioRxiv* (2020).
33. Shiozaki, H. M., Ohta, K. & Kazama, H. A Multi-regional Network Encoding Heading and Steering Maneuvers in Drosophila. *Neuron* **106,** 126–141.e5 (2020).
34. Hulse, B. K. *et al.* A connectome of the *Drosophila* central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife* **10** (ed Calabrese, R. L.) e66039 (2021).
35. Tang, S. & Guo, A. Choice Behavior of Drosophila Facing Contradictory Visual Cues. *Science* **294,** 1543–1547 (2001).
36. Guo, C. *et al.* A conditioned visual orientation requires the ellipsoid body in Drosophila. *Learning & Memory* **22,** 56–63 (2015).
37. Götz, K. G. Course-Control, Metabolism and Wing Interference During Ultralong Tethered Flight in Drosophila Melanogaster. *Journal of Experimental Biology* **128,** 35–46 (1987).
38. Reiser, M. B. & Dickinson, M. H. A modular display system for insect behavioral neuroscience. *Journal of Neuroscience Methods* **167,** 127 –139 (2008).
39. Brembs, B. & Heisenberg, M. Conditioning with compound stimuli in Drosophila melanogaster in the flight simulator. *J Exp Biol* **204,** 2849–2859 (2001).
40. Turner-Evans, D. B. *et al.* The Neuroanatomical Ultrastructure and Function of a Biological Ring Attractor. *Neuron* **108,** 145–163.e10 (2020).
41. Williams, C. B. Insect Migration. *Annual Review of Entomology* **2,** 163–180 (1957).
42. Coyne, J. A. *et al.* Long-Distance Migration of Drosophila. *The American Naturalist* **119,** 589–595 (1982).
43. Wehner, R. Astronavigation in Insects. *Annual Review of Entomology* **29,** 277–298 (1984).
44. Weir, P. T. & Dickinson, M. H. Flying Drosophila orient to sky polarization. *Curr Biol* **22,** 21–27 (2012).
45. Dickinson, M. H. Death Valley, Drosophila, and the Devonian Toolkit. *Annual Review of Entomology* **59.** PMID: 24160432, 51–72 (2014).
46. Leitch, K. J., Ponce, F. V., Dickson, W. B., van Breugel, F. & Dickinson, M. H. The long-distance flight behavior of Drosophila supports an agent-based model for wind-assisted dispersal in insects. *Proceedings of the National Academy of Sciences* **118** (2021).
47. Neuser, K., Triphan, T., Mronz, M., Poeck, B. & Strauss, R. Analysis of a spatial orientation memory in Drosophila. *Nature* **453,** 1244–1247 (2008).
48. Heisenberg, M., Wolf, R. & Brembs, B. Flexibility in a single behavioral variable of Drosophila. *Learn Mem* **8,** 1–10 (2001).
49. Dill, M., Wolf, R. & Heisenberg, M. Behavioral Analysis of Drosophila: Landmark Learning in the Flight Simulator. *Learning and Memory* **2,** 152–160 (1995).
50. Wolf, R. & Heisenberg, M. On the fine structure of yaw torque in visual flight orientation of Drosophila melanogaster. *J. Comp. Physiol. A* **140,** 69–80 (1980).
51. Muijres, F. T., Elzinga, M. J., Iwasaki, N. A. & Dickinson, M. H. Body saccades of Drosophila consist of stereotyped banked turns. *Journal of Experimental Biology* **218,** 864–875 (2015).
52. Warren, T. L., Weir, P. T. W. & Dickinson, M. H. Flying Drosophila melanogaster maintain arbitrary but stable headings relative to the angle of polarized light. *Journal of Experimental Biology* **221** (2018).
53. Haberkern, H. *et al.* Visually guided behavior and optogenetically induced learning in head-fixed flies exploring a virtual landscape. *Curr Biol* **29,** 1647–1659 (2019).
54. Mathejczyk, T. F. & Wernet, M. F. Heading choices of flying Drosophila under changing angles of polarized light. *Sci Rep* **9,** 16773 (2019).
55. Tang, S., Wolf, R., Xu, S. & Heisenberg, M. Visual Pattern Recognition in Drosophila Is Invariant for Retinal Position. *Science* **305,** 1020–1022 (2004).
56. Pan, Y. *et al.* Differential roles of the fan-shaped body and the ellipsoid body in Drosophila visual pattern memory. *Learning Memory* **16,** 289–295 (2009).
57. Tammero, L. & Dickinson, M. The influence of visual landscape on the free flight behavior of the fruit fly Drosophila melanogaster. *Journal of Experimental Biology* **205,** 327–343 (Feb. 2002).
58. Frye, M. A. & Dickinson, M. H. Visual Edge Orientation Shapes Free-Flight Behavior in Drosophila. *Fly* **1.** PMID: 18820449, 153–154 (2007).
59. Hartmann, G. & Wehner, R. The ant's path integration system: a neural architecture. *Biological Cybernetics* **73,** 483–497 (1995).
60. Wittmann, T. & Schwegler, H. Path integration — a network model. *Biological Cybernetics* **73,** 569–575 (1995).
61. Turner-Evans, D. B. *et al.* Angular velocity integration in a fly heading circuit. *eLife* **6,** e23496 (2017).
62. Stone, T. *et al.* An Anatomically Constrained Model for Path Integration in the Bee Brain. *Current Biology* **27,** 3069–3085.e11 (2017).

63. Cope, A. J., Sabo, C., Vasilaki, E., Barron, A. B. & Marshall, J. A. R. A computational model of the integration of landmarks and motion in the insect central complex. *PLOS ONE* **12,** 1–19 (2017).
64. Goulard, R., Buehlmann, C., Niven, J. E., Graham, P. & Webb, B. A unified mechanism for innate and learned visual landmark guidance in the insect central complex. *PLOS Computational Biology* **17,** 1–30 (Sept. 2021).
65. Lyu, C., Abbott, L. F. & Maimon, G. Building an allocentric travelling direction signal via vector computation. *Nature* **601,** 92–97 (2022).
66. Touretzky, D. S., Redish, A. D. & Wan, H. S. Neural representation of space Using sinusoidal arrays. *Neural Computation* **5,** 869–884 (Nov. 1993).
67. Matheson, A. M. *et al.* A neural circuit for wind-guided olfactory navigation. *bioRxiv* (2022).
68. Kim, S. S., Rouault, H., Druckmann, S. & Jayaraman, V. Ring attractor dynamics in the Drosophila central brain. *Science* **356,** 849–853 (2017).
69. Green, J. *et al.* A neural circuit architecture for angular integration in Drosophila. *Nature* **546,** 101–106 (2017).
70. Green, J. & Maimon, G. Building a heading signal from anatomically defined neuron types in the Drosophila central complex. *Current Opinion in Neurobiology* **52.** Systems Neuroscience, 156–164 (2018).
71. Seelig, J. & Jayaraman, V. Feature detection and orientation tuning in the Drosophila central complex. *Nature* **503,** 262–266 (2013).
72. Omoto, J. J. *et al.* Visual Input to the Drosophila Central Complex by Developmentally and Functionally Distinct Neuronal Populations. *Current Biology* **27,** 1098–1110 (2017).
73. Sun, Y. *et al.* Neural signatures of dynamic stimulus selection in Drosophila. *Nat Neurosci* **20,** 1104–1113 (2017).
74. Kottler, B., Faville, R., Bridi, J. C. & Hirth, F. Inverse control of turning behavior by dopamine D1 receptor signaling in columnar and ring neurons of the central complex in Drosophila. *Current Biology* **29,** 567–577.e6 (2019).
75. Frighetto, G., Zordan, M. A., Castiello, U., Megighian, A. & Martin, J.-R. Dopamine modulation of Drosophila ellipsoid body neurons, a nod to the mammalian basal ganglia. *Frontiers in Physiology* **13** (2022).
76. Beetz, M. J. *et al.* State-dependent egocentric and allocentric heading representation in the monarch butterfly brain. *bioRxiv* (2021).
77. Jacob, P.-Y. *et al.* An independent, landmark-dominated head-direction signal in dysgranular retrosplenial cortex. *Nature Neuroscience* **20,** 173–175 (2017).
78. Page, H. & Jeffery, K. Landmark-Based Updating of the Head Direction System by Retrosplenial Cortex: A Computational Model. *Frontiers in Cellular Neuroscience* **12,** 191 (July 2018).
79. Ofstad, T. A., Zuker, C. S. & Reiser, M. B. Visual place learning in Drosophila melanogaster. *Nature* **474,** 204–207 (2011).
80. Mizunami, M., Weibrecht, J. M. & Strausfeld, N. J. Mushroom bodies of the cockroach: their participation in place memory. *J Comp Neurol* **402,** 520–537 (1998).
81. Ardin, P., Peng, F., Mangan, M., Lagogiannis, K. & Webb, B. Using an Insect Mushroom Body Circuit to Encode Route Memory in Complex Natural Environments. *PLoS Comput Biol* **12,** e1004683 (2016).
82. Collett, M. & Collett, T. S. How does the insect central complex use mushroom body output for steering? *Current Biology* **28,** R733–R734 (2018).
83. Buehlmann, C. *et al.* Mushroom Bodies Are Required for Learned Visual Navigation, but Not for Innate Visual Behavior, in Ants. *Current Biology* **30,** 3438–3443.e2 (2020).
84. Sun, X., Yue, S. & Mangan, M. A decentralised neural model explaining optimal integration of navigational strategies in insects. *Elife* **9** (2020).
85. Kamhi, J. F., Barron, A. B. & Narendra, A. Vertical Lobes of the Mushroom Bodies Are Essential for View-Based Navigation in Australian Myrmecia Ants. *Curr Biol* **30,** 3432–3437 (2020).
86. Bennett, J. E. M., Philippides, A. & Nowotny, T. Learning with reinforcement prediction errors in a model of the Drosophila mushroom body. *Nat Commun* **12,** 2569 (2021).
87. Liu, Q. *et al.* Gap junction networks in mushroom bodies participate in visual learning and memory in Drosophila. *eLife* **5,** e13238 (2016).
88. Collett, T. S. & Land, M. F. Visual control of flight behaviour in the hoverfly Syritta pipiens L. *J. Comp. Physiol.* **99,** 1–66 (1975).
89. Mongeau, J.-M. & Frye, M. A. Drosophila Spatiotemporally Integrates Visual Signals to Control Saccades. *Current Biology* **27,** 2901–2914 (2017).
90. Yttri, E. A. & Dudman, J. T. Opponent and bidirectional control of movement velocity in the basal ganglia. *Nature* **533,** 402–406 (2016).
91. Zhou, B., Hofmann, D., Pinkoviezky, I., Sober, S. J. & Nemenman, I. Chance, long tails, and inference in a non-Gaussian, Bayesian theory of vocal learning in songbirds. *Proceedings of the National Academy of Sciences* **115,** E8538–E8546 (2018).
92. Jiang, W.-C., Xu, S. & Dudman, J. Construction of a hippocampal cognitive map depends upon spatial context. *Research Gate* (Jan. 2021).
93. Reichardt, W. & Wenking, H. Optical detection and fixation of objects by fixed flying flies. *Naturwissenschaften* **56,** 424–424 (Aug. 1969).
94. Wehner, R. Spontaneous pattern preferences of Drosophila melanogaster to black areas in various parts of the visual field. *Journal of Insect Physiology* **18,** 1531–1543 (1972).

95. Horn, E. The mechanism of object fixation and its relation to spontaneous pattern preferences in Drosophila melanogaster. *Biological Cybernetics* **31,** 145–158 (1978).
96. Maimon, G., Straw, A. D. & Dickinson, M. H. A simple vision-based algorithm for decision making in flying Drosophila. *Curr Biol* **18,** 464–470 (2008).
97. Grabowska, M. J. *et al.* Innate visual preferences and behavioral flexibility in Drosophila. *Journal of Experimental Biology* **221.** jeb185918 (Dec. 2018).
98. Panser, K. *et al.* Automatic Segmentation of Drosophila Neural Compartments Using GAL4 Expression Data Reveals Novel Visual Pathways. *Curr Biol* **26,** 1943–1954 (2016).
99. Klapoetke, N. C. *et al.* Ultra-selective looming detection from radial motion opponency. *Nature* **551,** 237–241 (2017).
100. Bolton, A. D. *et al.* Elements of a stochastic 3D prediction engine in larval zebrafish prey capture. *eLife* **8** (eds Berman, G. J., Calabrese, R. L., Washbourne, P. & Combes, S. A.) e51975 (2019).
101. Demir, M., Kadakia, N., Anderson, H. D., Clark, D. A. & Emonet, T. Walking *Drosophila* navigate complex plumes using stochastic decisions biased by the timing of odor encounters. *eLife* **9** (eds Seminara, A., Calabrese, R. L., Murthy, V. N. & Celani, A.) e57524 (2020).
102. Bishop, C. M. *Pattern Recognition and Machine Learning* Second (Springer-Verlag, 2006).
103. Botvinick, M. *et al.* Reinforcement Learning, Fast and Slow. *Trends in Cognitive Sciences* **23** (2019).
104. Wang, J. X. *et al.* Learning to reinforcement learn. *arXiv:1611.05763* (2016).
105. Duan, Y. *et al.* RL^2: Fast Reinforcement Learning via Slow Reinforcement Learning. *arXiv:1611.02779* (2016).
106. Harlow, H. The formation of learning sets. *Psychol Rev* **56** (1949).
107. Zador, A. M. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature Communications* **10,** 3770 (2019).
108. Guo, A *et al.* Conditioned visual flight orientation in Drosophila: dependence on age, practice, and diet. *Learning & Memory* **3,** 49–59 (1996).
109. JW, W., AM, W., J, F., LB, V. & R, A. Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. *Cell* **112,** 271–282 (2003).
110. Seelig, J. D. *et al.* Two-photon calcium imaging from head-fixed Drosophila during optomotor walking behavior. *Nature Methods* **7,** 535–540 (2010).
111. Hubert, M. & Vandervieren, E. An adjusted boxplot for skewed distributions. *Computational Statistics & Data Analysis* **52,** 5186–5201 (2008).
112. Shiozaki, H. & Kazama, H. Parallel encoding of recent visual experience and self-motion during navigation in Drosophila. *Nat Neurosci* **20,** 1395–1403 (2017).
113. B.j., H. *et al.* A visual pathway for skylight polarization processing in Drosophila. *eLife* **10,** e63225 (2021).
114. Heinze, S. & Reppert, S. M. Sun Compass Integration of Skylight Cues in Migratory Monarch Butterflies. *Neuron* **69,** 345–358 (2011).
115. Vitzthum, H., Müller, M. & Homberg, U. Neurons of the Central Complex of the Locust Schistocerca gregaria are Sensitive to Polarized Light. *Journal of Neuroscience* **22,** 1114–1125 (2002).
116. Okubo, T., Patella, P, D'Alessandro, I & Wilson, R. A Neural Network for Wind-Guided Compass Navigation. *Neuron* **107,** 924–940 (5 2020).
117. Hanesch, U., Fischbach, K.-F. & Heisenberg, M. Neuronal architecture of the central complex in Drosophila melanogaster. *Cell and Tissue Research* **257,** 343–366 (1989).
118. Isaacman-Beck, J. *et al.* SPARC enables genetic manipulation of precise proportions of cells. *Nature Neuroscience* **23,** 1168–1175 (2020).
119. Skaggs, W., Knierim, J., Kudrimoti, H. & McNaughton, B. A model of the neural basis of the rat's sense of direction. *Adv Neural Inf Process Syst* **7,** 173–180.
120. Kuntz, S., Poeck, B. & Strauss, R. Visual Working Memory Requires Permissive and Instructive NO/cGMP Signaling at Presynapses in the Drosophila Central Brain. *Current Biology* **27,** 613–623 (2017).
121. Liang, X. *et al.* Morning and Evening Circadian Pacemakers Independently Drive Premotor Centers via a Specific Dopamine Relay. *Neuron* **102,** 843–857.e4 (2019).
122. Franconville, R., Beron, C. & Jayaraman, V. Building a functional connectome of the *Drosophila* central complex. *eLife* **7** (eds VijayRaghavan, K, Scott, K. & Heinze, S.) e37017 (2018).
123. Scheffer, L. K. *et al.* A connectome and analysis of the adult *Drosophila* central brain. *eLife* **9** (eds Marder, E., Eisen, M. B., Pipkin, J. & Doe, C. Q.) e57443 (2020).
124. Lin, C. Y. *et al.* A comprehensive wiring diagram of the protocerebral bridge for visual information processing in the Drosophila brain. *Cell Rep* **3,** 1739–1753 (2013).
125. Wolff, T., Iyer, N. A. & Rubin, G. M. Neuroarchitecture and neuroanatomy of the Drosophila central complex: A GAL4-based dissection of protocerebral bridge neurons and circuits. *J Comp Neurol* **523,** 997–1037 (2015).
126. Currier, T. A., Matheson, A. M. & Nagel, K. I. Encoding and control of orientation to airflow by a set of Drosophila fan-shaped body neurons. *Elife* **9** (2020).
127. Lu, J. *et al.* Transforming representations of movement from body- to world-centric space. *Nature* **601,** 98–104 (2022).
128. Müller, J., Nawrot, M., Menzel, R. & Landgraf, T. A neural network model for familiarity and context learning during honeybee foraging flights. *Biological Cybernetics* **112,** 113–126 (2018).
129. Zhu, L., Mangan, M. & Webb, B. Spatio-temporal Memory for Navigation in a Mushroom Body Model. *bioRxiv* (2020).

130. Hu, W. *et al.* Fan-Shaped Body Neurons in the Drosophila Brain Regulate Both Innate and Conditioned Nociceptive Avoidance. *Cell Reports* **24,** 1573–1584 (2018).
131. Claridge-Chang, A. *et al.* Writing Memories with Light-Addressable Reinforcement Circuitry. *Cell* **139,** 405–415 (2009).
132. Aso, Y. & Rubin, G. M. Dopaminergic neurons write and update memories with cell-type-specific rules. *eLife* **5** (ed Luo, L.) e16135 (2016).
133. Cohn, R., Morantte, I. & Ruta, V. Coordinated and Compartmentalized Neuromodulation Shapes Sensory Processing in Drosophila. *Cell* **163,** 1742–1755 (2015).
134. Cognigni, P., Felsenberg, J. & Waddell, S. Do the right thing: neural network mechanisms of memory formation, expression and update in Drosophila. *Current Opinion in Neurobiology* **49.** Neurobiology of Behavior, 51–58 (2018).
135. Siju, K. *et al.* Valence and State-Dependent Population Coding in Dopaminergic Neurons in the Fly Mushroom Body. *Current Biology* **30,** 2104–2115.e4 (2020).
136. Liu, Q., Liu, S., Kodama, L., Driscoll, M. & Wu, M. Two Dopaminergic Neurons Signal to the Dorsal Fan-Shaped Body to Promote Wakefulness in ¡em¿Drosophila¡/em¿. *Current Biology* **22,** 2114–2123 (2012).
137. Ueno, T. *et al.* Identification of a dopamine pathway that regulates sleep and arousal in Drosophila. *Nature Neuroscience* **15,** 1516–1523 (2012).
138. Weir, P. T., Schnell, B. & Dickinson, M. H. Central complex neurons exhibit behaviorally gated responses to visual motion in Drosophila. *Journal of Neurophysiology* **111.** PMID: 24108792, 62–71 (2014).
139. Hsu, C. T. & Bhandawat, V. Organization of descending neurons in Drosophila melanogaster. *Scientific Reports* **6,** 20259 (2016).
140. Namiki, S., Dickinson, M. H., Wong, A. M., Korff, W. & Card, G. M. The functional organization of descending sensory-motor pathways in *Drosophila*. *eLife* **7** (ed Scott, K.) e34272 (2018).
141. Cande, J. *et al.* Optogenetic dissection of descending behavioral control in *Drosophila*. *eLife* **7** (ed Scott, K.) e34275 (2018).

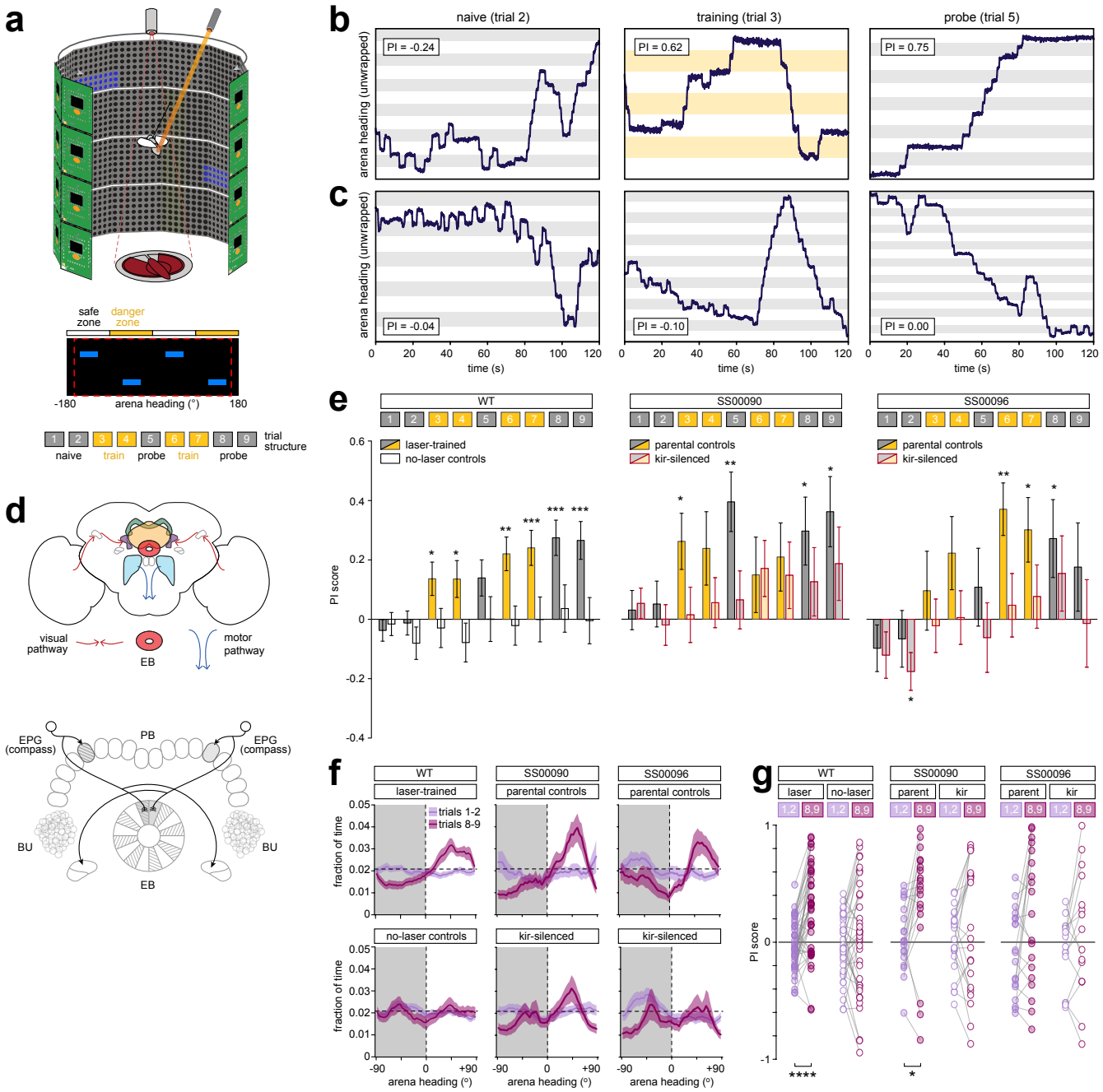
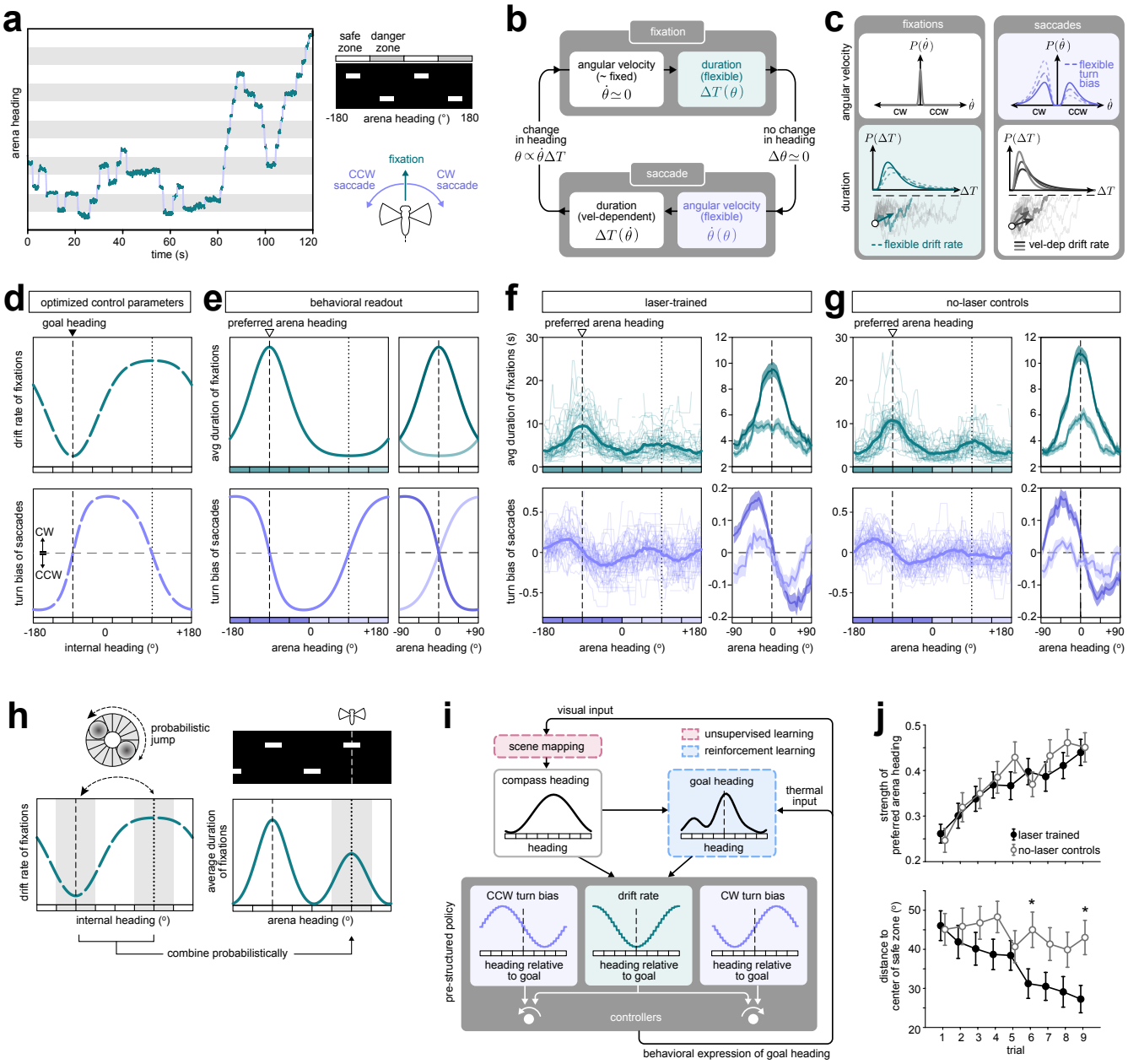

a
safe zone
danger zone
-180
arena heading (°)
180
trial structure
naive
train
probe
train
probe
b
naive (trial 2)
training (trial 3)
probe (trial 5)
PI = -0.24
PI = 0.62
PI = 0.75
arena heading (unwrapped)
c
PI = -0.04
PI = -0.10
PI = 0.00
time (s)
d
visual pathway
EB
motor pathway
EPG (compass)
PB
BU
e
WT
SS00090
SS00096
laser-trained
no-laser controls
parental controls
kir-silenced
PI score
f
trials 1-2
trials 8-9
fraction of time
arena heading (°)
g
laser
no-laser
parent
kir

Figure 1: Flies need their neural compass to operantly learn to avoid heat punishment in a 1D virtual visual environment. a) Upper: Schematic of flight simulator and LED arena. Middle: The arena is divided into four quadrants; one set of opposing quadrants contains horizontal bars at a high elevation, and the other set of quadrants contains horizontal bars at a low elevation. During training trials, one set of quadrants ('danger zone') is paired with an aversive heat punishment (orange bars), while the other set of quadrants ('safe zone') remain unpunished (white bars). The red dashed box indicates the span of the visual arena (see Methods). Lower: training protocol. During training trials, laser punishment is delivered whenever the fly's heading falls within the danger zone. During naive and probe trials, no punishment is delivered. Each trial lasts 2 min. **b)** Example heading trajectories and performance indices (PI scores) from a single fly that underwent laser training. Trajectories were unwrapped to reveal the overall structure in the behavior (Methods). Gray and yellow bars indicate headings within the danger zone that are punished (yellow bars; training trial), or that will be/have been punished (gray bars; naive/probe trials). Flies exhibit periods of straight flight in which they maintain a constant arena heading (fixations), punctuated by abrupt saccades that lead to changes in heading (saccades). **c)** Same as **b**, for a no-laser control fly that did not undergo laser training. **d)** Upper: Schematic of the fly central complex (CX), highlighting visual pathways to the ellipsoid body (EB) and motor pathways leaving the CX. Lower: Schematic illustrating simplified representation of EPG (compass) neurons that maintain an internal representation of the fly's head direction. Shown are two compass neurons, each innervating a single 'wedge' of the EB and a single 'glomerulus' of the protocerebral bridge (PB). The population of compass neurons tiles the entire EB and PB. **e-g)** Comparison of learning performance for three different genetic backgrounds of flies ('WTB', 'SS00090', and 'SS00096') that underwent laser training ('laser-trained' and 'parental controls'). These are compared with three genetically-matched groups in which flies did not undergo laser training ('no-laser controls'; WTB) or underwent laser training but with Kir expressed in EPGs ('kir-silenced'; SS00090 and SS00096). **e)** Average PI scores across trials. Error bars: mean +/- standard error. Significance: two-sided Wilcoxon signed rank test (*$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$) against the null hypothesis that the PI scores come from a distribution with a median of 0. **f)** Average residency (expressed as a fraction of total time) spent at different headings in the arena, measured before (light pink) and after (dark pink) training and aligned to the safe/danger zones. Note that arena headings are folded from $\pm 180°$ to $\pm 90°$ based on the symmetry of the visual scene. Error bars: mean +/- standard error. **g)** Changes in PI scores before and after laser training. Significance: paired, two-sided Wilcoxon signed rank test (*$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; ****$p \leq 0.0001$) against the null hypothesis that the difference in PI scores between naive and probe trials has a median of 0. Numbers in colored boxes at top indicate trials used to compute PI scores.


a
arena heading
time (s)
safe zone
danger zone
-180
arena heading (°)
180
fixation
CCW saccade
CW saccade
b
fixation
angular velocity (~ fixed)
duration (flexible)
change in heading
no change in heading
saccade
duration (vel-dependent)
angular velocity (flexible)
c
fixations
saccades
angular velocity
duration
flexible turn bias
flexible drift rate
vel-dep drift rate
d
optimized control parameters
goal heading
drift rate of fixations
turn bias of saccades
CW
CCW
internal heading (°)
e
behavioral readout
preferred arena heading
avg duration of fixations
arena heading (°)
f
laser-trained
g
no-laser controls

h
probabilistic jump
drift rate of fixations
internal heading (°)
average duration of fixations
arena heading (°)
combine probabilistically

i
visual input
scene mapping
unsupervised learning
reinforcement learning
compass heading
goal heading
thermal input
heading
pre-structured policy
CCW turn bias
drift rate
CW turn bias
heading relative to goal
controllers
behavioral expression of goal heading

j
strength of preferred arena heading
laser trained
no-laser controls
distance to center of safe zone (°)
trial

Figure 2: An inferred behavioral policy captures conditioned and unconditioned behavior. a) Left: same behavioral trace as shown in the left panel of Fig 1b, segmented into fixations (green) and saccades (purple), as schematized at bottom right. **b)** A behavioral policy to generate alternating modes of fixations and saccades in model flies. Each mode is approximated by sampling a fixed angular velocity $\dot{\theta}$ that is maintained over a duration of time ΔT (see panel **(c)**). Boxes shaded green and purple highlight parameters that we hypothesize are flexibly modified as a function of heading. **c)** Left: fixations are approximated by periods of straight flight that are maintained at near-zero angular velocity (upper box) but that last a variable duration of time (lower box). The observed variability in fixation durations is captured by a drift diffusion process with flexible drift rates, whereby higher drift rates lead to shorter events (Methods). Right: saccades are approximated by periods of non-zero angular velocity (upper box) that are maintained over a velocity-dependent duration of time (lower box). The observed variability in saccades is captured by sampling an angular velocity from one of two lognormal distributions whose relative weights are controlled by a flexible directional bias, combined with a drift diffusion process with a velocity-dependent drift rate that controls the duration of saccades. We assume that both the drift rate of fixations and the directional bias of saccades are controlled as a function of the fly's internal representation of heading θ. **d-e)** When optimized to maintain a single goal heading (indicated by the dashed vertical line), the optimal control parameters are structured as a function of the fly's current heading relative to the goal heading. When the fly is at the goal heading, a low drift rate leads to long fixations (**d-e**, upper panels). When the fly is away from the goal heading, a high drift rate leads to short fixations, and a directional saccade bias drives the internal heading toward the goal heading (**d-e**, lower panels); note that a CW saccade will move the fly CW but the internal heading will shift CCW, and hence the lower panels in **d** and **e** are mirror-symmetric versions of one another. For both fixations and saccades, the optimal control parameters generate qualitatively different behavior about the goal (dashed line) and 'anti-goal' (dotted line) headings; the right column of **e** highlights this comparison for $180°$ of heading space about the goal heading (dark shaded boxes in left column and dark line in right column) and anti-goal heading (light shaded boxes in left column and light line in right column). **f-g)** Left columns: average duration of fixations (upper panels) and direction of saccades (lower panels) measured in laser-trained (**f**; $n = 44$) and no-laser control flies (**g**; $n = 40$), aligned to the inferred goal headings of individual flies (vertical dashed line; vertical dotted line marks the anti-goal heading). Thin lines mark data for individual flies; thick lines mark population averages. Right columns: comparison of behavioral patterns observed about the goal headings (dark lines) and anti-goal headings (light lines), plotted in the same style as panel **e**. **h)** We hypothesize that when the control parameters shown in **d** are coupled to an unstable internal representation of heading (left column), the resulting behavioral readout is bimodal (right column). This bimodality arises because the internal representation can jump between orientations corresponding to symmetric views of the visual scene, and thus leads to a 'copying-over' of the behavioral policy at symmetric arena headings. **i)** Hypothesized architecture of 'pre-structured' behavioral policy (gray box) that controls the duration of fixations and the directionality of saccades as a function of the relative difference between the fly's compass and goal headings. The behavior that is generated from this policy determines the visual and thermal inputs that the model fly receives. The visual input is mapped onto an internal representation of heading via an unsupervised learning process (red box), which in turn determines how the compass heading tethers to different views of the visual scene [29, 30, 63]. The compass heading is, in turn, used to update an internal goal heading based on thermal reinforcement (blue box). The two learning systems—one unsupervised, the other reinforced—together determine how the fly's internal goal heading is behaviorally expressed through the pre-structured policy. **j)** The strengths of the preferred arena headings of both laser-trained and no-laser control flies strengthen over time (upper); in laser-trained flies, these headings additionally shift away from danger and toward safety (lower). Significance: two-sided Wilcoxon rank sum test (*$p \leq 0.05$) against the null hypothesis that quantities measured for laser-trained and no-laser control flies come from continuous distributions with the same medians.

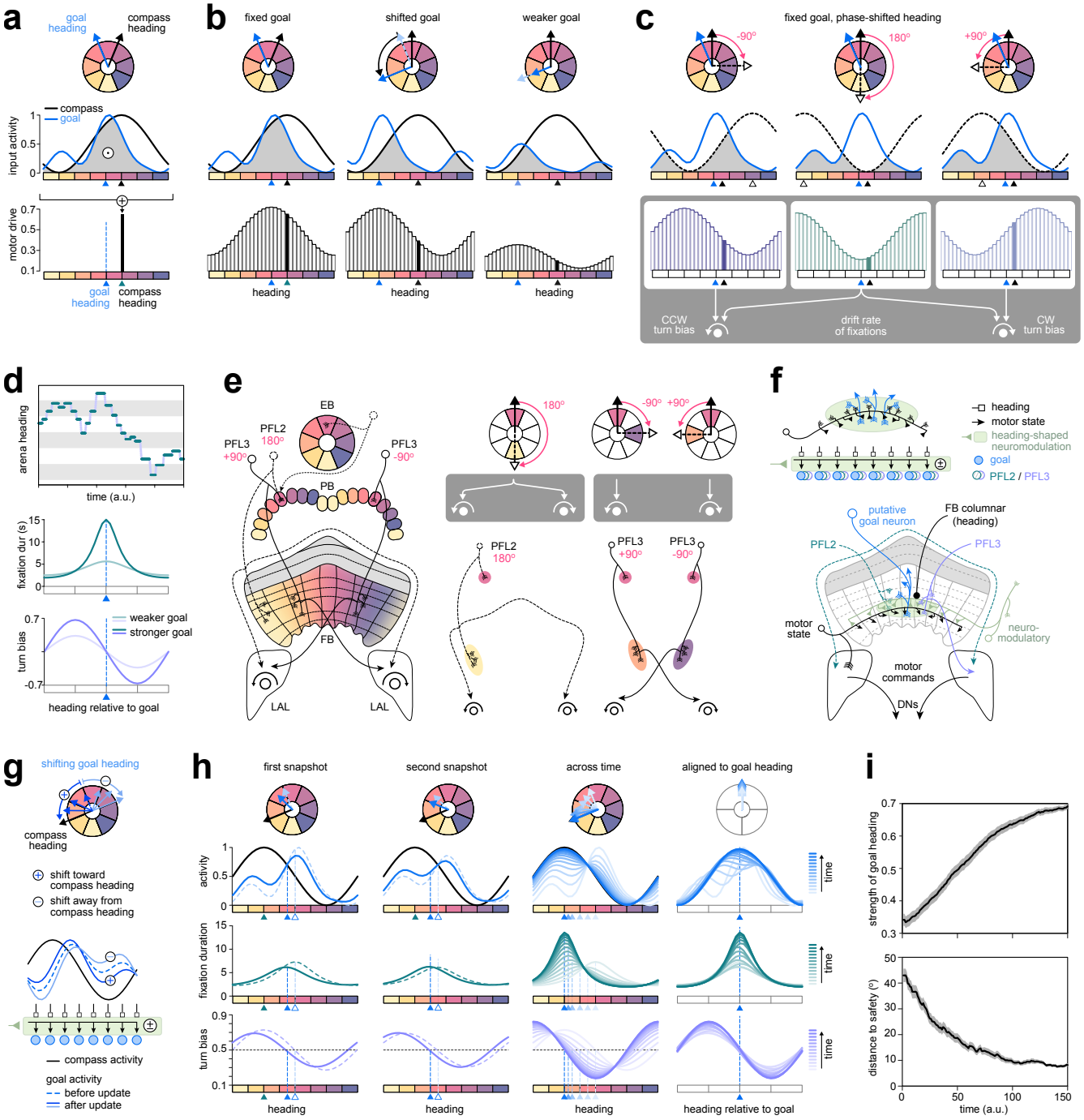

a
goal heading
compass heading
compass
goal
input activity
motor drive
goal heading
compass heading
b
fixed goal
shifted goal
weaker goal
heading
c
fixed goal, phase-shifted heading
-90°
180°
+90°
CCW turn bias
drift rate of fixations
CW turn bias
d
arena heading
time (a.u.)
fixation dur (s)
turn bias
weaker goal
stronger goal
heading relative to goal
e
EB
PFL2 180°
PFL3 +90°
PFL3 -90°
PB
FB
LAL
LAL
180°
-90° +90°
PFL2 180°
PFL3 +90°
PFL3 -90°
f
heading
motor state
heading-shaped neuromodulation
goal
PFL2 / PFL3
putative goal neuron
FB columnar (heading)
PFL2
PFL3
motor state
neuro-modulatory
motor commands
DNs
g
shifting goal heading
compass heading
shift toward compass heading
shift away from compass heading
compass activity
goal activity
before update
after update
h
first snapshot
second snapshot
across time
aligned to goal heading
activity
fixation duration
turn bias
time
heading
heading relative to goal
i
strength of goal heading
distance to safety (°)
time (a.u.)

Figure 3: Hypothesized neural implementation of a pre-structured behavioral policy tethered to a flexible goal heading. a) Schematic of multiplication of a sinusoidal population vector representing heading activity (black curve) and a population goal vector of arbitrary shape (blue curve) to generate motor drive (black bar). The multiplicative operation is similar to computing the overlap between the two activity profiles (gray shaded region). Phasor representation (top) illustrates the angular orientation, or phase (colored wedges), of the compass and goal headings; this is also indicated by the black and blue triangles below the activity profiles. We used the circular mean of the arbitrarily-shaped goal vector to represent its phase. **b)** First column: for a fixed goal profile (top row, blue arrow; middle row, blue curve), a specific compass heading (black arrow/curve) will generate a particular motor drive (bottom row, black filled bar). Sweeping across all compass headings reveals a structured, sinusoidal motor drive that varies as a function of the phase difference between the compass and goal heading (histogram). Second and third columns: Shifting the goal profile (blue curve) for a fixed compass heading shifts the motor drive while retaining its overall structure (histogram); similarly, weakening the goal profile reduces the motor drive. The multiplicative operation ensures that the motor drive remains sinusoidal, peaked at the goal heading and structured as a function of the phase difference between compass and goal headings, regardless of their absolute orientations. **c)** Shifting the compass heading by phase shifts of $-90°$, $+180°$, and $+90°$ results in motor drives that are shifted by $+90°$, $-180°$, and $-90°$, respectively. If these are used to drive downstream controllers, we recover the architecture of the pre-structured policy schematized in Fig 2i. **d)** The policy in **c** generates behavior that qualitatively mimics what we observe in real flies (compare to Fig 2a), with longer fixations at the goal heading (middle row) and directed saccades toward the goal heading on average (bottom row). Illustrated for a weak and strong goal heading. **e)** Left: The anatomical projection patterns of PFL 'action neurons' may implement the phase shifts and premotor projections schematized in **c**. Colors representing phases were propagated from EB wedges to PB glomeruli based on compass neuron projection patterns (see Fig 1d), and to columns of another CX region, the fan-shaped body (FB), based on the morphology of 'zero-phase-shift' PB-FB columnar neurons that project from PB glomeruli with approximately the same phase to an overlapping columnar region in the FB [34, 62, 65]. Note that phases shown are a simplification of true phases. Center: The PFL2 neuron population implements a $+180°$ phase shift between the PB and FB, and projects bilaterally to two premotor areas (the lateral accessory lobes, LALs) that drive CW and CCW saccades. The phase shift can be seen by noting how one PFL2 neuron projects from the pink glomerulus in the PB to yellow columns in the FB. All PFL2 neurons display similar phase shift motifs and together tile the FB, covering all angles. Right: two populations of PFL3 neurons implement $-90°$ and $+90°$ phase shifts and project unilaterally to the LAL, which drives either CW or CCW saccades. These phase shifts can be seen by noting how a neuron from one PFL3 population projects from the pink glomerulus in the PB to the orange columns in the FB, and then projects unilaterally to the LAL region controlling CW saccades. A neuron from the other PFL3 population projects from the pink glomerulus in the PB to the purple columns in the FB, and then unilaterally to the LAL region controlling CCW saccades. All left and right PFL3 neurons display phase shift motifs similar to the two sample neurons shown here. Together, the populations tile the FB and cover all angles. **f)** Hypothesized columnar neurons in the FB that putatively store the goal heading in synaptic weights. In our model, these weights are modified in a heading- and reinforcement-dependent manner by tangential dopaminergic neurons that innervate layers of the FB and receive local, heading-tuned synaptic input, as suggested by the CX connectome [34]. **g)** Schematic of dopamine-mediated Hebbian plasticity acts on goal weights carried by synapses between tangential motor-state neurons and FB columnar 'goal neurons'. Dopaminergic reinforcement is heading-dependent and is shaped locally by a heading signal conveyed by (other) FB columnar neurons (not shown). Positive reinforcement drives the goal heading toward the compass heading (dark blue arrows), while negative reinforcement drives the goal heading away from the compass heading (light blue arrows). **h)** Illustration of the evolution of the goal profile (blue arrow/curves, first/second rows) when the model fly is at a fixed compass heading (black arrows/curves, first/second rows) and experiencing positive reinforcement. Over time (first three columns), the goal profile is strengthened at the compass heading, and weakened away from the compass heading (compare dashed (before) and solid (after) blue curves in first/second columns), leading to a shift in the goal heading (compare open and filled blue triangles in first/second columns). The behavioral readout (third/fourth rows) shifts with the goal heading. If we align different temporal snapshots to the goal heading (fourth column), we see that the behavioral readout is always aligned to the goal heading, even as the goal heading is shifting over time. As the goal heading becomes stronger, with a larger amplitude and more sinusoidal shape (darker blue curves), the behavioral readout also strengthens (darker green/purple curves). **i)** The learning rule illustrated in **g**, when coupled with the policy in **c**, causes the goal heading to strengthen over time and shift toward safety, mimicking what we observe in real flies (see Fig 2j).

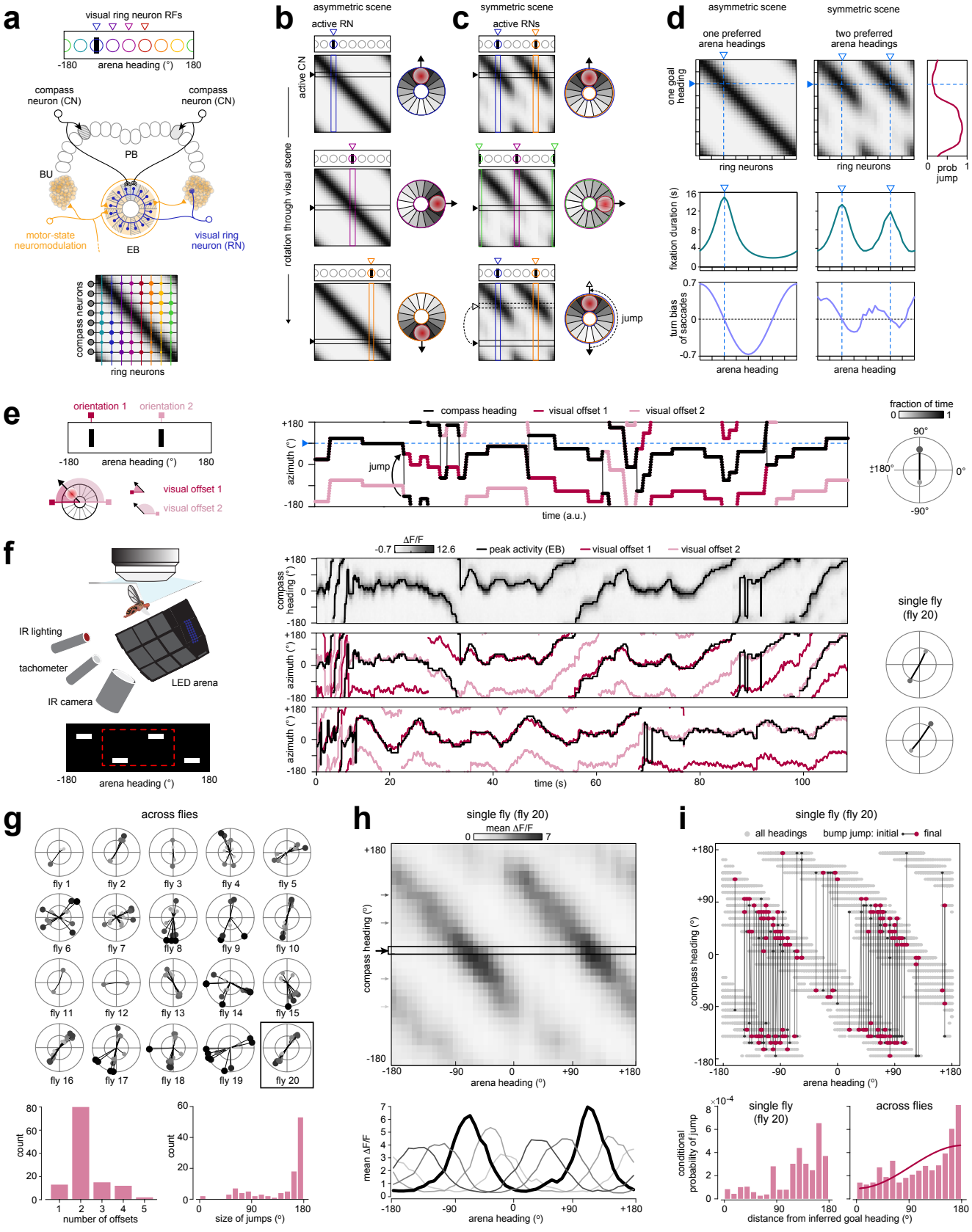

a
visual ring neuron RFs
-180
arena heading (°)
180
compass neuron (CN)
compass neuron (CN)
PB
BU
motor-state neuromodulation
EB
visual ring neuron (RN)
compass neurons
ring neurons
b
asymmetric scene
active RN
active CN
rotation through visual scene
c
symmetric scene
active RNs
jump
d
asymmetric scene
symmetric scene
one preferred arena headings
two preferred arena headings
one goal heading
ring neurons
prob jump
fixation duration (s)
turn bias of saccades
arena heading
e
orientation 1
orientation 2
visual offset 1
visual offset 2
compass heading
azimuth (°)
jump
time (a.u.)
fraction of time
90°
±180°
0°
-90°
f
IR lighting
tachometer
IR camera
LED arena
ΔF/F
-0.7
12.6
peak activity (EB)
compass heading (°)
time (s)
single fly (fly 20)
g
across flies
fly 1
fly 2
fly 3
fly 4
fly 5
fly 6
fly 7
fly 8
fly 9
fly 10
fly 11
fly 12
fly 13
fly 14
fly 15
fly 16
fly 17
fly 18
fly 19
fly 20
count
number of offsets
size of jumps (°)
h
single fly (fly 20)
mean ΔF/F
compass heading (°)
arena heading (°)
i
all headings
bump jump: initial
final
conditional probability of jump
distance from inferred goal heading (°)

Figure 4: Mapping of symmetric visual scene onto heading representation leads to structured instability in compass neuron dynamics. **a)** Top row: visual ring neurons are activated by visual features (for example, here, a vertical stripe) appearing on the arena screen (circles, color corresponds to arena orientation) within their spatial receptive fields (denoted by triangles). Anatomical schematic in second row: visual ring neurons (blue lines) make all-to-all inhibitory synapses (filled blue circles) onto compass neurons (black) that are idealized as only receiving input in the EB and sending outputs to the PB. Neuromodulatory neurons whose activity carries the fly's motor state —in our model, the fly's angular velocity— project throughout the EB and BU. Bottom row, left: Heatmap illustrates the strength of total ring neuron inhibition onto compass neurons; synaptic strength of individual ring neuron connections onto different compass neurons is shown by the radius of filled colored circles (different colors correspond to different ring neurons). Bottom row, right: inhibitory Hebbian-like plasticity weakens synapses based on co-activity between (in this case, blue) ring and compass neurons during saccades [29]. o/p: output; i/p: input. **b)** In an asymmetric scene with a single visual pattern, the behavioral policy shown in Fig 3 stabilizes a self-consistent mapping between ring and compass neurons (left column, indicated by the diagonal structure in the weight matrix). This mapping ensures that both visual scene shifts and angular velocity inputs caused by saccades accurately update the heading representation (right column shows bump in EB); ring neurons that are activated by a particular view of the visual scene (black bar in small circle) weakly inhibit (dark color in the EB) compass neurons with the appropriate heading tuning and more strongly inhibit (lighter color) compass neurons with different heading tuning, thereby ensuring that the compass neuron activity bump remains tethered to movements of the scene. **c)** In a symmetric scene with two repeating visual patterns, the behavioral policy stabilizes a two-fold symmetric mapping (denoted by repeating light and dark diagonal bands in the connection weight matrix) in which two arena headings opposite each other map onto the same internal compass heading. This mapping results in some compass neuron (sample row indicated by black filled triangle) will be weakly inhibited (dark color in mapping) and others strongly inhibited (light color in mapping) by multiple ring neurons of $180°$-opposite spatial receptive fields (orange and blue empty triangles in top row). Top row: The bump currently occupies an EB location in which the compass neurons are weakly inhibited (see position of bump in EB schematic at right). Middle row: When the fly saccades, a different set of compass and ring neurons are activated (compare triangles in first and second rows; note shift in position of visual stimuli in vertical columns representing the arena). Bottom row: A further saccade results in the compass bump moving into an EB location in which the activated compass neurons receive strong inhibitory synaptic input from the activated ring neurons (third row; bump would have been in the light part of the mapping matrix and the EB). This would likely cause the bump to jump to a different location in the EB, where these same active ring neurons only weakly inhibit compass neurons (darker color in mapping); in a two-fold symmetric scene, this corresponds to a jump of $180°$. **d)** Left column: The behavioral readout (second and third rows) resulting from the mapping in an asymmetric scene is unimodal and centered on the preferred arena heading (empty blue triangle and vertical dashed line), which corresponds uniquely to an internal goal heading (filled blue triangle and horizontal dashed line in mapping). Right column: Because the policy drives the fly and its compass heading toward the internal goal heading (filled blue triangle and dashed horizontal line), the compass bump's movements will, on average, tend to be in the neighborhood of the goal heading more than away from it. This, in turn, weakens the inhibition from active ring neurons onto compass neurons whose tuning is near the goal heading (indicated by the alignment of the dark structures in the weight matrix with the goal heading. Thus, the two-fold symmetry would lead the bump to jump most frequently when it is farthest from the internal goal heading (red probability curves to the right of the mappings). These bump dynamics result in a bimodal behavioral readout (second and third rows). The strength of the mapping, the probability of bump jumps, and the behavioral readout all depend on the strength of the goal heading. **e)** Left: we tracked the orientation of the compass heading relative to different angular 'offsets' of the visual scene (pink square markers). For a scene with two repeating visual patterns, there are two offsets (measured as the circular distance between the peak activity and a reference orientation of the visual scene) that correspond to visually-identical views of the scene. Right: as a model fly navigates with respect to the scene shown in **e**, the compass heading (black trace) tracks one view of the scene (pink trace) for some time, and then jumps to the symmetric view of the scene (red trace). Far right: Offsets for the simulated trial. Size, color, and radial distance of marker from the origin indicate fraction of time spent at a given offset. **f)** Left: schematic of two-photon calcium imaging setup for tethered flies flying within an LED arena. Right: First row: EPG neuron calcium activity (heatmap) and location of peak activity (black line) during closed-loop tethered flight with a symmetric visual scene. Second and third rows: the peak activity (black) jumps between two different offsets (pink and maroon) that correspond to symmetric views of the scene. Shown for trials 1-2. Far right: Offsets for the two trials. **g)** Top: Polar plots showing offsets for all trials for each fly. Bottom row, left: Number of unique offsets, aggregated across flies and trials. Right: Angular size of bump jumps, aggregated across flies and trials. **h)** Main panel: Average $\Delta F/F$ of different wedges in the EB (compass heading) as a function of arena heading, averaged across all 9 trials for the fly shown in panel **f**. Lower panel: Tuning curve for different EB wedges (curve in bold corresponds to the wedge denoted by an arrow in the main panel). **i)** Bump jumps as a function of location in the EB, collapsed across all 9 trials for the fly in (**f**), shown for jumps that exceeded a threshold of $135°$. Lower: Probability of a bump jump as a function of angular distance from the preferred bump heading, shown for a single fly (left panel) and aggregated across flies (right panel). For each fly, the conditional probability was estimated as the fraction of all visits to a given EB wedge (accumulated across trials) for which the angular size of the bump jump exceeded $135°$. Bars: average probability across flies. Solid line: best-fitting cosine curve.

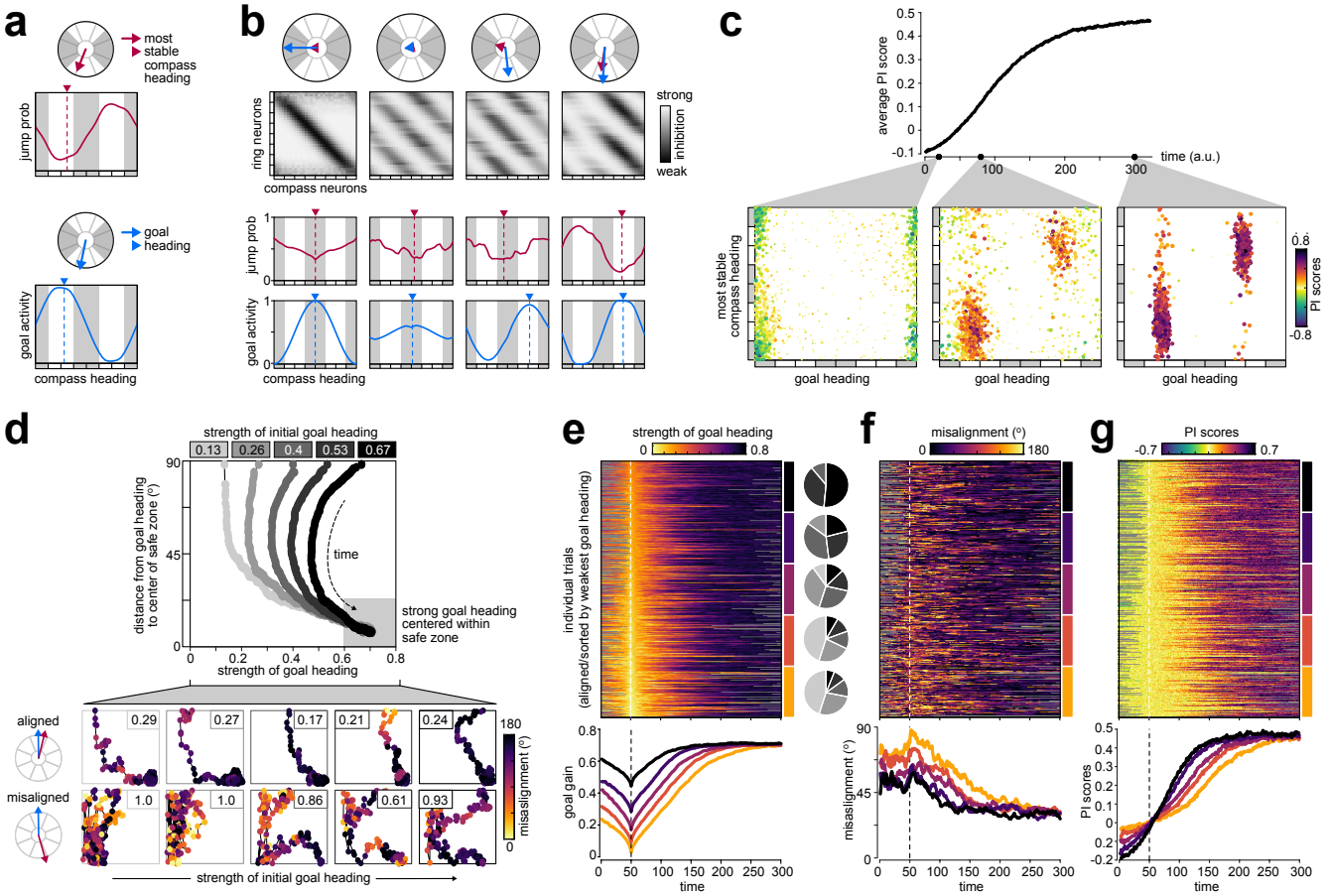


Figure 5: Interaction between visual scene mapping and goal learning shapes individual variability. a) Schematic of two learning systems in the context of the learning paradigm outlined in Fig 1a. We track the most stable heading (upper) and the goal heading (lower) relative to the safe and danger zones (white and gray regions, respectively). **b)** Example learning trajectory from a single model fly that was initialized with a stable visual map derived from an asymmetric scene (heatmap in second row) and a strong goal vector aligned with the center of the danger zone (blue arrow in top row, blue curve in bottom row). Over time (successive columns), in response to punishment, the goal vector weakens and begins to shift toward the safe zone. In parallel, different views of the visual scene induce plasticity in the same subsets of ring neurons, which lead to the development of symmetries in the mapping of the visual scene and shape the probability that the bump will jump to and from different orientations (change in red curves in third row). Over time, the most stable compass heading and the goal heading align with each other, and with the center of the safe zone (top row, bottom two rows). **c)** Upper panel: average PI scores over time, averaged over 1000 model flies. As in panel **b**, each model fly was initialized with a stable visual map derived from an asymmetric visual scene, and with a goal heading that was aligned with the center of the danger zone but that differed in strength. Lower panels: over time, the most stable headings align with the goal headings, and both align with the safe zone. **d)** Upper panel: average properties of the goal heading as it evolves over time, averaged over groups of model flies that began with the same initial strength of goal headings. Consistent with panel **c**, model flies eventually shift and strengthen their goal headings within the center of the safe zone, reaching the gray box in the lower right corner. Lower panels: individual model flies vary substantially in their trajectories through the same space. Individual columns highlight flies that began with increasingly strong goal vectors; top and bottom rows indicate flies that were fast or slow to learn, respectively. Numbers in each box indicate the fraction of total simulation time required for each fly to reach the lower right corner highlighted in gray above. Trajectories are colored by the degree of alignment between the most stable heading and the goal heading; brighter colors indicate less alignment. Flies that were slow to learn (bottom row) suffered from prolonged periods of poor alignment. **e-g)** Upper panels: properties of learning trajectories over time, aligned by the time when the goal heading was the weakest (vertical dashed lines) and sorted by the value of the weakest goal heading. Solid rectangles in upper panels indicate the groupings that were used to construct the averages shown in the lower panels. Pie charts indicate the fraction of model flies within each group that began with the goal strengths shown in **d**. Lower panels: same as upper panels, but averaged over groups of flies that exhibited similar dynamics in their goal headings. **e)** Model flies that began with strong initial goal headings were more likely to maintain strong goal headings throughout training; note that the top group (purple), which consists of flies that began with strong goal headings (primarily dark wedges in pie chart), shows only modest drops in their goal headings over time, whereas the bottom group (orange), which consists of flies that began with weak goal headings (primarily light wedges in pie chart), shows large drops in their goal headings over time. **f)** During period where the goal heading is weak (bright regions in **e**), model flies tend to suffer periods of misalignment, where their visual maps become misaligned with their goal headings. **g)** Periods of misalignment (bright regions in **f**) can slow the learning process, leading to prolonged periods of low PI scores.

METHODS

Experimental Methods

Fly culture. Parental flies were grown sparsely on Wurzburg food in bottles for at least 6 generations [108]. Crosses were first done in vials then transferred to bottles after 1-3 days, followed by transferring to a new bottle every day to limit F1 density. 10 males and 25 virgins were used for each cross. The day after eclosion, F1s were transferred to a new bottle with a piece of kimwipe for self-cleaning and transferred again to a new bottle with kimwipe the day before imaging or behavioral experiments. All experiments were performed with 5-6 days old female flies.

The crosses for the flies used in the experiments are: WT: 11D03AD males x WTB virgins (Figures 1, 2) EPG silencing: SS males x WTB;;UAS-Kir2.1-EGFP virgins (Figure 1) Parental controls: SS males x WTB virgins (Figure 1) EPG two-photon imaging: 60D05 males x 20xUAS-GCaMP6f [su(Hw) attP5; attP2, VK00005] (WTB) (Figure 4)

Visual arena. A blue LED circular arena [38] was assembled with 44 panels (4 rows and 11 columns, spanning 120° in elevation and 330° in azimuth), with the LED emission peaking at 464 nm (Bright LED Electronics Corp., BM-10B88MD). Two layers of blue filter (Roscolux #59) were laid on top of the LED panels to allow 0.04% transmission. Each fly was tethered at the end of a tungsten wire and positioned in the center of the arena. An 880 nm LED (Digi-Key, PDI-E803-ND) illuminated the fly from above. A custom-built wingbeat analyzer (University of Chicago Electronics Shop) measured the wingbeat frequency and amplitudes for both wings. Yaw turning was computed as the left minus right wingbeat amplitude. A computer (Dell, R5500) controlled the timing of the experiments through a data acquisition device (National Instruments, USB-6229 BNC) and sampled the flight parameters at 1 kHz. This is in contrast to most well-established visual learning studies, which have relied on torque meters to measure the fly's instantaneous torque and drive the rotation of a paper drum imprinted with visual patterns [18, 48].

Flight visual learning. The 360° yaw space around the fly was divided into 4 quadrants. A single horizontal blue bar (37.5°w x 11.25°h) was displayed in each quadrant, with alternating elevations at +/- 30°, such that the pattern repeats every 180°. Throughout the assay, the fly had closed-loop control of the visual pattern it was flying towards. The unconditioned stimulus (US) as punishment was a fiber-coupled infrared laser (Edmund Optics, 975 nm, 400 mW) modulated by a 10 kHz square wave with varying duty cycles output from a function generator (Agilent, 33220A, 20 MHz) and gated by the specific positions of the arena pattern such that either the higher bar quadrant or the lower bar quadrant was accompanied by the laser punishment aimed at the back of the fly. The laser was turned off during the pre-training naïve trials and post-training memory/probe trials. The visual pattern was jump-rotated randomly to a new position after every trial. A 100 ms air puff towards the fly was triggered whenever the fly stopped flying. However, only data during flight from flies that flew continuously without stop or puffing for more than 60 s in all 3 trial types were included for further analysis. All visual stimulation and behavior parameters were recorded with a data acquisition device (National Instruments, USB-6229 BNC). During no-laser mock experiments, the US laser was not turned on. The EPG silencing and parental control experiments were performed double-blind by RK.

Fly preparation for imaging during flight. Flies were transferred to a polypropylene tube using a custom 3D-printed funnel positioned on the top of the opened bottle, then anaesthetized in a custom brass cold plate at 4°C. The largest female was selected to fit onto a custom aluminum mounting bridge cooled to 4°C, and held down with vacuum suction ventrally. The bridge was then rotated to hold the fly upside down and an inverted custom laser-milled PEEK holder pushed up the fly's head from below, with a center hole lined up under the head. Small drops of UV-activated epoxy were used to glue the fly head, thorax and the back of the head capsule to the holder. Another small drop was used to glue the proboscis. The eyes were kept completely below the holder to allow unhindered visual stimulation and most of the back head plate was exposed through the center hole in the holder. The legs were left intact and the wings kept free to flap during flight because of the inverted pyramid shape of the holder. The back plane of the head was angled at approximately 26° to match the angle of the visual arena under the two-photon microscope. For imaging experiments, we used an LED arena with 18 panels (3 rows and 6 columns, spanning 90° in elevation and 180° in azimuth). For flight experiments, only flies that could fly continuously for 90 s while maintaining closed-loop stripe fixation after mounting were selected. Artificial hemolymph as described previously [109] was used to fill the holder reservoir from the top. A window was carefully opened on the back head

capsule with a tungsten dissection probe and fine forceps and the trachea underneath were gently picked away to allow optic access to the brain [110].

Two-photon calcium imaging. Calcium imaging was performed on a two-photon microscope (Bruker Nano, formerly Prairie Technologies). A Chameleon Vision II or Discovery laser (Coherent) tuned to 920 nm was used with the power adjusted to the lowest sufficient level, usually between 3 and 20 mW at the sample. A resonant galvanometer mirror was used to scan the laser beam along the x-axis at 8 kHz, resulting in a frame rate of 60 Hz with 256 by 256 resolution. For volume imaging, a piezo motor drove the 40x objective (Olympus, LUMPlanFl/IR, NA 0.8) along the z-axis. The 2-plane z-stack acquisition was repeated over time throughout the trial at a rate of 14.5 volumes/s. The green and red channel signals, when applicable, were collected through a set of dichroic mirror (575 nm) and band-pass filters (525 + 35 nm for green, 607 + 22.5 nm for red). A GaAsP photomultiplier tube (Hamamatsu, 7422PA-40) was used to acquire data from each channel. Each imaging series was triggered from the experiment-controlling computer.

Data Analysis

All data analysis was performed in MATLAB (MathWorks Inc., Natick, MA).

Partitioning behavior into fixations and saccades. All data analyses were performed after segmenting behavioral traces into fixations and saccades. We developed a custom algorithm to perform this segmentation (SI Fig S2a). We first filtered the difference in wingbeat amplitude between left and right wings, A^{WB}, using a bandpass filter of order 6, with a lower cutoff frequency of $0.1\ \mathrm{Hz}$ and an upper cutoff frequency of $10\ \mathrm{Hz}$ (we'll denote this filtered signal as $\tilde{A}^{\mathrm{WB}}$). We then used sign changes in the filtered amplitude to segment the trajectory into a set of individual turns; each turn in this set was thus defined as a sequence of time points $\{t\}$ for which $\tilde{A}^{\mathrm{WB}}_{\{t\}}$ had a consistent sign.

A turn that produces a sustained nonzero difference in wingbeat amplitude $\tilde{A}^{\mathrm{WB}}$ will lead to changes in the arena heading x in the opposite direction. We used this to define a quantity $s_t = -\tilde{A}^{\mathrm{WB}}_t|\Delta x_t|$, where $\Delta x_t = x_{t+1} - x_t$ is the instantaneous change in arena heading (allowing for wrapping between pixel 96 and pixel 1). This quantity measures the coherence between differences in wingbeat amplitude and changes in arena heading; s_t will be large in magnitude during times when changes in wingbeat amplitude lead to large and coherent changes in arena heading, and will be zero when changes in wingbeat amplitude do not lead to a change in arena heading. We thus used this signal to select turns that fall into the former category, where s is large in magnitude.

To this end, we first selected candidate saccades as those turns that led to a total change in arena heading of at least 2 pixels ($7.5°$). For this subset of turns, we used s_t to refine the beginning and end of individual turns. We defined the beginning of the turn as the timepoint t_{start} for which there was the largest instantaneous change $\Delta s_t = s_{t+1} - s_t$, and the end of the turn as the first timepoint thereafter for which s_t dropped below $1/4$ of its maximum value (i.e., $t_{\mathrm{end}} : s_t < \frac{1}{4} s_{t_{\mathrm{start}}}$). The remaining timepoints ($t < t_{\mathrm{start}}$, and $t > t_{\mathrm{end}}$) were segmented as separate turns. We repeated this process until all large turns had been refined in this way.

This resulted in a refined set of turns; these turns included both the candidate saccades that led to a change in arena heading, and the small turns that did not lead to a change in arena heading. We removed all turns that during which either (i) the wingbeat frequency f^{WB} dropped below a threshold of $f^{\mathrm{WB}}_{\mathrm{min}} = 0.001$ (for the upright arena) and $f^{\mathrm{WB}}_{\mathrm{min}} = 1$ (for the two-photon arena) for any amount of time, or (ii) the turn intersected with a period of time that spanned $100\ \mathrm{ms}$ before and $500\ \mathrm{ms}$ after an airpuff. We then ranked each remaining turn according to a quantity $r(\mathrm{turn}) = |\langle s_{t_{\mathrm{start}}:t_{\mathrm{end}}}\rangle(x_{t_{\mathrm{end}}} - x_{t_{\mathrm{start}}})|$ that combines the average change in wingbeat amplitude $\langle s_{t_{\mathrm{start}}:t_{\mathrm{end}}}\rangle$ with the total change in arena heading $(x_{t_{\mathrm{end}}} - x_{t_{\mathrm{start}}})$. This quantity will be largest for turns that are large and fast, which comprise a small fraction of the entire set of turns. We thus used an outlier detection procedure to identify those turns for which r exceeded a threshold $r_{\mathrm{thresh}} = Q3 + \exp(3M) * 1.5 * \mathrm{IQR}$. Here, $Q3$ is the third quartile (or 75%) of r, IQR is the inter-quartile range, and M is a skewness estimated using the med-couple of r [111]. r_{thresh} was estimated separately for individual flies and trials, based on the distributions of turns produced by the given fly in the given trial.

The set of turns for which $r > r_{\mathrm{thresh}}$ were classified as saccades. The periods of time between each saccade were then further segmented and classified as either fixations or periods of drift. For each of these periods of time between two saccades, we used the MATLAB function *findchangepts.m* to detect changes in the mean and slope of the arena heading x (using the linear statistic in *findchangepts.m* and a minimum threshold on the residual error of 100000); this results in a set of candidate segments to classify as fixations of drift. We again removed any segments that intersected with an airpuff event; we also removed any segments for which f^{WB} dropped below $f^{\mathrm{WB}}_{\mathrm{min}}$ for more

than $30\mathrm{ms}$. For each remaining segment, we determined the slope of the best fitting line; any segments that had a slope greater than $0.003\ \mathrm{pixels/ms}$ were classified as periods of drift; the remaining segments were classified as fixations.

In the main text, we focused our analysis on fixations and saccades. We described these two behavioral modes in terms of their duration and their angular velocity; here, the angular velocity (in degrees/ms) is given by $\omega = -50(360/96)\tilde{A}^{\mathrm{WB}}$, where the factor 50 converts between wingbeat amplitude and pixels/ms, and the factor $(360/96)$ converts from pixels/ms to degrees/ms. Saccades were characterized by short durations and large average angular velocities, while fixations were characterized by long durations and low average angular velocities (SI Fig S2b).

Characterizing fixation properties. Individual fixations varied substantially in their duration, and the distribution of these durations was heavy-tailed. We therefore considered three putative heavy-tailed distributions: log-normal, inverse Gaussian, and generalized Pareto. We fit each of these three distributions to the distribution of fixation durations $P(\Delta t)$ under two different conditions: when fixations were accumulated across flies within a given trial, and separately when fixations were accumulated across trials for a given fly. We performed this fitting for laser-trained and no-laser control flies.

Prior to fitting, we removed fixations whose durations were below a variable threshold $\Delta t_{\mathrm{thresh}}$. We then evaluated fitting performance for 25 evenly spaced values of $\Delta t_{\mathrm{thresh}}$ between 20 and 500 ms. We used the MATLAB function *fitdist.m* to perform the fitting, and we used the Bayesian information criterion (BIC) to evaluate fits. We found that the inverse Gaussian distribution, $\mathrm{IG}(\Delta t; \mu, \lambda)$, was the best-fitting distribution across a majority of scenarios (trials or flies) for thresholds between 100 and 300 ms; within this range, a threshold of 200 ms maximized this number of scenarios for which the inverse Gaussian was the best fit. We therefore performed the remainder of our analysis on fixations whose duration exceeded $\Delta t_{\mathrm{thresh}} = 200$ ms.

The inverse Gaussian distribution is characterized by two parameters: a mean μ and a shape parameter λ. This distribution can be generated by a drift diffusion to bound process with a mean drift rate ν, spread η^2, and bound a (SI Fig S2g). This process yields an inverse Gaussian distribution $P(\Delta t) = \mathrm{IG}(\Delta t; a/\nu, a^2/\eta^2)$ whose parameters $\mu = a/\nu$ and $\lambda = a^2/\eta^2$ are defined in terms of the parameters of the diffusion process. When we compared the best-fitting values of μ_F and λ_F across different datasets (where the subscript F denotes that parameters were fit to the distribution of fixations), we found that the variability in these parameters was consistent with a drift diffusion process with a variable drift rate ν_F but fixed spread η_F^2 and bound a_F. To illustrate this, note that the mean $\mathrm{mean} = \mu$ and variance $\mathrm{var} = \mu^3/\lambda$ of the inverse Gaussian distribution satisfy $\log(\mathrm{var}) = 3\log(\mathrm{mean}) - \lambda$. If the variability in the fit parameters can be explained by changes in ν alone (with fixed η^2 and a), the plot of $\log(\mathrm{var})$ versus $3\log(\mathrm{mean})$ will be well-described by a line of slope 1 and fixed offset $-\lambda = -a^2/\eta^2$. SI Fig S2h shows this comparison when the mean and variance are computed from the fit parameters (filled markers) versus estimated directly from the data (open markers), along with a line of slope of 1 and best-fitting offset $\lambda_F = 0.63$ (dashed line). We found that this provided a better fit than a model in which the bound is variable, and the drift rate and spread are fixed (dotted line).

We used this result to posit that fixations are controlled by a drift diffusion process with a fixed spread η_F and bound a_F, but an adaptive drift rate ν_F. Because there are three parameters of the drift diffusion process but only two parameters needed to define the inverse Gaussian distribution, we are free to choose one of the drift diffusion parameters and fit the other two. We chose to set $\eta_F = 1$, which requires that $a_F = 0.79$ (thus satisfying $\lambda_F = a_F^2/\eta_F^2 = 0.63$). When we restricted our analysis to fixations that were initiated within the danger zone during the first 60s of the first training trial (and remained within the danger zone for 95% of their duration), we found that they were well fit by the same process, but with a higher drift rate and thus shorter average duration (red star in SI Fig S2h). The reduction in fixation duration in response to heat can thus be captured by an additional "reflexive" drift process with drift rate $\nu_F = 0.38$, spread $\eta_F = 1$, and bound $a_F = 0.79$.

Characterizing saccade properties. Individual saccades varied in both their speed and duration. We found that the average duration of saccades depended on their average angular speed; we thus began by characterizing the distribution of angular speeds $P(\omega)$, where ω is computed by averaging the instantaneous difference in wingbeat amplitude over the duration of a saccade. The angular change in heading can be computed from this via $\Delta\theta = \omega\Delta t$. We then characterized the distribution of durations conditioned on speed, $P(\Delta t|\omega)$.

Prior to fitting, we removed saccades whose speeds were below a threshold $\omega_{\mathrm{thresh}} = 0.1$. We used the MATLAB function *allfitdist.m* to fit 16 different parametric distributions to the distribution of speeds $P(\omega)$ under two different conditions: when saccades were accumulated across flies within a given trial, and separately when saccades were accumulated across trials for a given fly. We performed this fitting for both laser-trained and no-laser control flies,

and we used BIC to evaluate fits. We found that the lognormal distribution $\mathrm{logn}(\Omega; \varphi, \sigma^2)$, with location φ and scale σ^2, was the best-fitting distribution across the majority of conditions. We then computed the directional bias in saccades, measured as $(N_{\mathrm{CW}} - N_{\mathrm{CCW}})/(N_{\mathrm{CW}} + N_{\mathrm{CCW}})$, where N_{CW} and N_{CCW} respectively denote the total number of clockwise and counter clockwise saccades taken within a single trial. We found that this bias also varied across trials (SI Fig S2d upper), suggesting that flies can adaptively control directional bias of their saccades.

The distribution $P(\omega)$ specifies the probability of initiating saccades of different speeds. For saccades of a given speed ω, there is significant variability in their duration (SI Fig S2f). To characterize this variability, we considered 36 equally spaced values of ω between 0.25 and 2. For each value of ω, we used the MATLAB function *allfitdist.m* to determine the parametric function that best fit the distribution of saccade durations $P(\Delta t|\omega)$ accumulated across flies, trials, and datasets. We found that these distributions were best fit by an inverse Gaussian distribution with fixed spread $\eta_S = 1$ and bound $a_S = 0.57$ but variable drift rate ν_S (SI Fig S2i), analogous to fixations. In this case, the drift rate increased nonlinearly with the speed $|\omega|$ (inset of SI Fig S2i); we used a least-squares fit to determine the parameters of the best-fitting sigmoid $f(|\omega|) = f_M/[1 + \exp(-k(|\omega| - \omega_0))] - f_0$; these where given by $f_M = 2.6$, $k = 7.36$, $\omega_0 = 0.32$, and $f_0 = -1.01$. Thus, for a saccade initiated with speed ω, the duration can be generated via a drift diffusion process with a velocity-dependent drift rate $\nu_S = f(|\omega|)$, and fixed values of $\eta_S = 1$ and $a_S = 0.57$.

Inferring the structure of a behavioral policy. Together, the analysis of fixations and saccades enable us to construct a behavioral policy that accounts for variability in the initiation, speed, and duration of both fixations and saccades (SI Fig S2j). Each behavioral mode (fixation versus saccade) is generated via a sequence of three steps: (i) Select an angular velocity by sampling from a parametrized distribution. For fixations, we approximate the angular velocity to be zero. For saccades, we sample the magnitude of the angular velocity from a lognormal distribution, and we take the directional bias (corresponding to the likelihood of initiating a clockwise versus counter clockwise saccade, which specifies the sign of the angular velocity) to be an adaptive parameter. (ii) Generate the duration via an online drift diffusion process with a variable drift rate. For fixations, we take this drift to be an adaptive parameter. For saccades, this drift is determined by the angular velocity selected in step (i). (iii) Determine the resulting change in heading, which is proportional to the product between the average angular velocity and the duration.

In the main text, we considered a simplification of this full behavioral policy in which the angular size of saccades (measured in deg) was directly sampled from a lognormal distribution with parameters $\varphi_S = 3.89$ and $\sigma_S = 0.54$ (fit to the distribution of saccade sizes across all flies, trials, and datasets; this generates saccades with an median angular size of $49°$), and assuming a fixed saccade duration of $t_S = 320$ ms (equal to the median saccade duration measured across all flies, trials, and datasets). Table 1 summarizes these choices.

Data selection. For all analyses shown in the main text, we used only those trials for which the fly exhibited at least $30\mathrm{s}$ of continuous flight (these periods of continous flight were defined using the set of saccades, fixations, and drift that we extracted from our segmentation; see *Analysis Methods: Partitioning behavior into fixations and saccades*). We kept all additional periods of flight beyond this length, regardless of length. This data was used in its entirety to compute PI scores. When analyzing fixations and saccades, we further excluded periods of drift; we then selected those fixations that exceeded a duration of $\Delta t_{\mathrm{thresh}} = 200$ms, and those saccades that exceeded an average angular velocity of $\omega_{\mathrm{thresh}} = 0.1$/ms.

Residency. Fig 1f shows the fraction of time spent in different regions of the arena, computed for data that was aggregated across trials 1-2 and trials 8-9. We computed this with respect to individual pixels in the arena as a function of their location relative to the safe and danger zones. Due to the two-fold symmetry of the visual scene, this resulted in pairs of pixels that shared the same location relative to safety/danger. For each fly, we summed the time (measured in ms) spent at each such pair of pixels, and we divided this by the total number of timepoints that met the selection criteria described above. Fig 1b shows the average and standard error of this fraction of time, measured across flies, for each such location in the arena. The dashed horizontal line shows the value that would be expected if flies spent equal amounts of time at all locations within the arena, given by $1/48 = 0.0208$.

Performance scores. Fig 1e shows the performance index (PI) scores, averaged across flies on each trial; Fig 1g shows PI scores for data aggregated across trials 1-2 and trials 8-9. PI scores were computed as $\mathrm{PI} = (T_{\mathrm{safe}} - T_{\mathrm{danger}})/(T_{\mathrm{safe}} + T_{\mathrm{danger}})$, where T_{safe} and T_{danger} denote the total time spent in safe and danger zones, respectively [18].

Aligning to individual preference. Fig 2f-g shows the average duration of fixations and directionality of saccades after aligning the behavioral data to the arena preference of each fly within each trial. To perform this alignment, we first computed the fraction of time that the fly resided at each of 96 orientations (corresponding to 96 pixel locations) within the arena. We then constructed an idealized residency profile that took a peak value of one at a central set of two pixels, and decayed linearly to zero over a span of 24 pixels in either direction (CW and CCW). We shifted this idealized profile with respect to the true residency profile of the fly, and we identified the preferred orientation as the one that maximized the overlap between the idealized and true residency profiles.

Computing heading-dependent averages. To perform the heading-dependent averages shown in Fig 2f-g, we first selected the set of saccades and fixations taken by each fly within probe trials 1, 2, 5, 8 and 9. We binned saccades according to the arena heading at which they were initiated; we binned fixations according to the average heading computed across the duration of the fixation. We used overlapping bins of width 15 pixels, centered on a given pixel. We then computed the average direction of saccades initiated within each bin, and similarly the average duration of fixations within each bin. For the data shown in Fig 2f-g, we included only those bins that had at least 2 samples per fly (either 2 fixations or 2 saccades per fly), among those, only those bins for which we had data from at least 5 flies; for these bins, we displayed both the individual averages per fly, and the population average across flies (thin and thick lines in the left column of 2f-g, respectively). We then displayed the mean and standard error in each of these quantities for the two quadrants centered about the preferred arena heading, and the two quadrants centered about the anti-preferred arena head (dark and light lines in the right column of 2f-g, respectively)

Measuring the strength of the preferred arena heading. The upper panel of Fig 2j shows the strength of the preferred arena heading, averaged across flies. To compute this, we first aligned the behavioral data to the preferred arena heading for individual flies on individual trials, as described in *Analysis Methods: Aligning to individual preference*. We then computed the overlap of the arena residency (measured as the fraction of total time spent at each heading) with the sawtooth profile shown in the inset of Fig 2j. This overlap measures how coherently the behavior was structured within the quadrants corresponding to the preferred and anti-preferred arena headings, and penalizes residency in the other two quadrants. Fig 2j shows the average and standard error of this quantity for laser-trained and no-laser control flies.

Measuring distance to safety. The lower panel of Fig 2j shows the distance from the preferred arena heading to safety, averaged across flies. To compute this, we first aligned the behavioral data to the preferred arena heading for individual flies on individual trials, as described in *Analysis Methods: Aligning to individual preference*. We then computed the minimum angular distance between this preferred heading and the center of the closest safe zone. Fig 2j shows the average and standard error of this distance for laser-trained and no-laser control flies.

Computing calcium transients. For volume imaging of EPG GCaMP activity (Fig 4f), we used two z-planes that together captured the dorsal and ventral halves of the EB. The image stack at each time step was converted into a summed intensity projection that was used for further analysis. We manually divided the EB into 32 wedge-shaped ROIs to capture population EPG activity in the structure. An additional ROI without any EPG arborization and outside the EB was selected to estimate background signal, including from leaked LED arena light. Time series of GCaMP activity for all EB ROIs were obtained by taking the average of the fluorescence signal within each ROI at each time step. The calcium transient for each ROI, $\Delta F/F_0$, was computed by subtracting fluorescence in the background ROI from all other ROIs, and using the lowest 10th percentile of background-subtracted fluorescence from each ROI as F0. The resulting time series were filtered using a simple boxcar (moving mean) filter (width 344 ms).

Rather than compute the population vector average (PVA), as in past work [19, 29, 40, 61, 68], we focused here on tracking peaks in EPG population activity ('bump position' or 'bump heading') at each time step. This allowed us to to track offsets between the EB location of the bump and the position of the visual scene at every time point, and to easily visualize changes in offsets, bump jumps, as seen in Fig 4f. Considering the symmetry of the visual scene, we tracked the position of the visual scene using two $180°$-offset time series, with the first being shifted by the first offset and the second by a second offset (if present, see below).

Clustering bump offsets. Fig 4f-g shows the offsets between the bump and the visual scene for individual flies. To cluster these offsets, we used the MATLAB function *kde.m* (with 256 mesh points) to perform kernel density estimation of the distribution of offsets for each fly on each trial. We then used the MATLAB function *findpeaks.m*

to determine the peaks in this density; we used offset values corresponding to these peaks as our candidate offset values. We then used the same function to determine the minima in this density (using a peak threshold of 10^{-4}), and we used the offset values corresponding to these minima as the bounds between different clusters. We the computed the sum of the density function within these bounds, divided by the sum of the density function over all time, and used this as the fraction of time spent at each offset. Fig 4f-g shows the fraction of time spent at different offsets for individual flies on individual trials.

Characterizing the number and angular separation of offsets. Fig 4g shows the total number of and angular separation between offsets. To construct these histograms, we first computed the fraction of time that the HD bump spent at different offsets relative to the visual scene for each fly on each trial, as described above. We then computed the number of instances (aggregated over flies and trials) that we observed a given number of distinct offsets; these results are shown in the lower left panel Fig 4g. For flies that exhibited two or more offsets on a given trial, we compute the angular distance between the dominant two offsets; this histogram is shown in the lower right panel of Fig 4g.

Computing HD tuning curves. Fig 4h shows the tuning of EB wedges to different arena headings. We first determined all times (aggregated across all 9 trials) that the visual scene was oriented at a particular angle relative to the fly, and then computed the average fluorescence transients $\Delta F/F$ of each wedge for each given scene orientation (see *Analysis Methods: Computing calcium transients*). The HD tuning curves below the main panel of Fig 4h show the average tuning of individual wedges as a function of the fly's heading in the arena, i.e., the "arena heading", which differs from the scene orientation by a sign flip.

Determining the locations of bump jumps with the EB. Fig 4i shows bump jumps as a function of their location within the EB. To determine the location of the bump jumps, we first determined the relationship between the arena heading and bump phase that minimized the angular distance between successive time points; an example of this relationship is shown in the main panel of Fig 4i. We took advantage of our previous results (shown in the lower right panel of Fig 4g) to select those changes in bump phase between $135°$ and $225°$; this range captured the majority of the bump jumps in our data. For each jump, we marked the location within the EB at which the jump was initiated. We then computed the angular distance from this location to the location of the putative preferred bump heading within the EB. To determine this location, we used the behavioral data for the same fly on the same trial to infer a preferred arena heading, as described above. For a given preferred arena heading, we determined the corresponding location in the EB at which the heading bump spent the most time. We used this as the location of the putative preferred bump heading in the EB. The lower left panel of Fig 4i shows the number of jumps that were initiated at a given angular distance from this goal location, divided by the total number of times that the bump visited locations of the same angular distance, for a single fly (accumulated across trials). The lower right panel of Fig 4i shows the same histogram, now accumulated across flies. To approximate these jump statistics, we determined the parameters of the best-fitting cosine function that minimized the mean-squared error between the measured and fit values of the histogram.

Note that the behavioral experiments were performed in arenas with a $330°$ angular span, but the imaging experiments were performed in arenas with a $180°$ span in the azimuth. Although we cannot rule out the possibility that the reduced horizontal span of the visual scene in imaging experiments affected the probability of the EPG bump jumping, similar bump instabilities have been reported in both flying and walking flies in symmetric visual settings in larger arenas as well [19, 29, 30].

Modeling

Determining the optimal policy for maintaining a goal heading. Fig 2d-e shows the drift rate and average duration of fixations, and the turn bias of saccades, that result from training a flexible RL agent to exhibit a preference for a goal heading. The learning algorithm is described in detail in *SI: Reinforcement learning framework* (see Algorithm S4); the parameters used in the model are summarized in Table 1.

Briefly, we learned a single set of weights $\vec{\omega} = \{\vec{\omega}_F, \vec{\omega}_S\}$ that specify the properties of fixations ($\vec{\omega}_F$) and saccades ($\vec{\omega}_S$) as a function of angular orientation (via a set of 16 von Mises radial basis functions), and we reported the resulting behavior when averaged over 100 different training runs. Reinforcement was delivered as a function of the current arena heading relative to a preferred heading; we assumed this reinforcement decayed linearly away from the preferred heading (see *SI: Reinforcement learning framework* for more details).

Prior to each training run, we initialized the set of flexible policy parameters $\vec{\omega}_F = 0.1$ and $\vec{\omega}_S = 0$, and we randomly initialized the arena heading to one of 96 evenly-spaced values between 0 and $360°$. Following each training run, we used Eq. 17 to evaluate the drift rate of fixations, $\nu(\theta; \vec{\omega})$, and the probability of rightward saccades, $p_R(\theta; \vec{\omega})$, as a function of bump heading θ given the learned parameters $\vec{\omega}$. We then computed the average duration of fixations $a_F/\nu(\theta; \vec{\omega})$, and the average turn bias of saccades $2p_R(\theta; \vec{\omega}) - 1$. We averaged these across training runs to produce the curves shown in Fig 2d-e.

Fig 2h illustrates the expected behavioral readout if we couple the optimal policy to an unstable internal representation of heading. To illustrate this, we assumed that the internal heading could jump between orientations that correspond to symmetric views of the visual scene; for the two-fold symmetric scene used here, this corresponds to a jump of $180°$. We further assumed that the jumps occurred probabilistically, and were least likely to occur at the preferred heading and most likely to occur at the symmetric (or "anti-preferred") heading. We used a cosine function to parametrize this probability.

Circuit model summary. In Figs 3-5, we constructed and simulated a circuit model that could implement and modify the parameters of a pre-structured behavioral policy. This model is described in detail in *SI: Reinforcement learning framework*. Briefly, we model a fly that can fixate and saccade. The duration of its fixations and the directionality of its saccades are determined by three populations of action neurons that receive phase-shifted input about the fly's current heading (from a population of compass neurons) and input about the fly's goal heading (from a population of goal neurons). Both the duration of fixations and the directionality of saccades are modulated by the strength of the goal heading, and by the fly's current bump heading relative to this goal heading. The location and strength of the goal heading are determined by a set of plastic goal weights $\vec{\omega}_G$ that can change over time based on the fly's current bump heading and current level of reinforcement. We additionally account for the stability of the heading itself, which is dictated by a set of plastic weights $\vec{\omega}_C$ from inhibitory ring neurons onto compass neurons. These weights are updated during saccades based on the velocity of the saccade, the fly's current heading during the saccade, and the current view of the visual scene [29]. We use these weights to determine the probability that the HD bump will jump between locations that correspond to symmetric views of the visual scene (see Eq. 27 in *SI: Reinforcement learning framework*). Bump jumps were assumed to occur immediately following a saccade. Both sets of weights were updated using simple Hebbian-like plasticity rules (see Equations 38-39 in *SI: Reinforcement learning framework*). We assume that the compass weights, $\vec{\omega}_C$, can be modified continuously, regardless of the presence or absence of reward/punishment. In contrast, we assume that the goal weights, $\vec{\omega}_G$, can only be modified in the presence of reward/punishment. Algorithms 3-4 show how we implemented this model.

To illustrate the behavior of the circuit model, as shown in Figs 3-5, we partitioned the space of arena and compass headings into $N = 32$ evenly-spaced values between $-180°$ and $180°$ (below, we will index these headings as $\theta_i, i \in [1, N]$). To mimic the experimental setup, we simulated two safe zones and two danger zones, each spanning $90°$ of heading space. We defined the centers of the safe zone to be at the arena headings $\theta_A = \{90°, 270°\} = \{\theta_9, \theta_{25}\}$. We used this model to simulate a period of training in which the compass and/or goal weights were evolving over time, and to probe performance and behavior given a fixed set of compass and goal weights. Each training period consisted of a series of iterations, each consisting of a single saccade and a single fixation. During simulations in which the goal weights were changing over time, we subsampled periods of fixation into 100ms increments, and iteratively updated the weights for each increment. Weights were not updated during saccades. During simulations in which the compass weights were changing over time, we iteratively updated the weights at each angular increment (each value of θ_i) during the saccade. Weights were not updated during fixations.

The duration of fixations and directionality of saccades were determined by the current goal weights, as described above (see Eq. 35 in *SI: Reinforcement learning framework* for more details), and the sizes of saccades were sampled from a lognormal distribution with parameters $\varphi_S = 3.89$ and $\sigma_S = 0.54$ (matched to the values that were fit from data). We assumed that all saccades lasted a fixed duration of $300\ \mathrm{ms}$ (taken from the median duration observed in data). During probe periods in which the goal weights remained fixed, there was no simulated reward or punishment. During training periods, the model fly received a reward of $+1$ per unit time when in the safe zone, and a reward of -1 per unit time when in the danger zone.

A pre-structured behavioral policy tethered to a flexible goal heading. Fig 3a-c illustrates the behavioral policy whose heading-dependent structure is guaranteed by a multiplicative operation between the compass and goal activity profiles. We used a goal activity profile of $\omega_G(\theta) = \vec{\omega}_0 \equiv \cos^2(\theta - \theta_3) + \cos^3(\theta - \theta_5)$, normalized to the range $[0, 1]$. We assumed a cosine profile of compass activity, also normalized to the range $[0, 1]$. The motor drive was determined by multiplying the compass and goal profiles and summing the output. We repeated this

calculation for each possible circular shift of the compass activity to compute the net output as a function of current relative to goal heading (Fig 3a-b). When computing this for phase-shifted compass headings (Fig 3c), the phase shift was applied before the multiplication with the goal activity profile. Fig 3d illustrates an example behavioral trajectory generated by this policy using a goal profile of $\omega_G(\theta) = g\cos(\theta - \theta_8) + 1/2$, where $g = [0.25, 0.5]$ for a weak and strong goal, respectively. We assumed that the compass weights were fixed and uniquely specified the arena heading. The average fixation duration and turn bias, shown in the lower panels of Fig 3c, were computed via Eqs. 33-35.

Fig 3g-h illustrates the temporal evolution of the fixation duration and turn bias as we modify the goal heading via Hebbian plasticity. The Hebbian-like learning update that we used is given in Eq. 39, and example weight updates are shown in Fig 3g for positive and negative reinforcement at a fixed compass heading. Fig 3h illustrates a continual updating of the goal weights given positive reinforcement at a fixed compass heading, and it illustrates the corresponding fixation durations and turn biases for each update (again computed via Eqs. 33-35).

Fig 3i illustrates the evolution of the goal vector in a simulated environment that mimics the behavioral task. We simulated 100 model flies with randomly initialized goal headings (all chosen to have a cosine profile that was peaked at the goal heading, but that varied in the orientation of the goal heading and the strength of the profile about that goal heading). As the goal heading evolved over time in response to reinforcement from the simulated environment, we tracked the circular mean (and its distance to the center of the closest safe zone) and strength of the goal profile. Fig 3i shows the mean and standard error of these quantities, computed across model flies.

Flexible mapping of visual scenes onto the compass heading. Fig 4a-e illustrates a model of anti-Hebbian plasticity that modifies synaptic weights from ring neurons onto compass neurons. We simulated the evolution of this weight matrix for 32 ring neurons and 32 compass neurons. This partitioned the space of both arena headings and compass headings into angular units of $360°/32 = 11.25°$). Compass neurons were assumed to maintain a von Mises bump profile that was normalized to the range $[0, 1]$, with concentration parameters $\kappa = \pi$, and whose location faithfully tracked changes in heading generated by the behavioral policy described above. Ring neurons were assumed to uniformly tile visual space with a receptive field width of three angular units (i.e., $33.75°$), such that two adjacent receptive fields had an overlap of one angular unit. For asymmetric visual scenes (Fig 4b), we assumed that each ring neuron fired at a maximum rate of 1 whenever a fixed orientation of the visual scene aligned with the center of its receptive field, and fired at half of its maximum rate whenever the orientation of the visual scene was shifted by one angular unit to either side of its receptive field center. For symmetric visual scenes (Fig 4c-e), we assumed that the ring neuron exhibited these same firing patterns but with respect to two symmetric orientations of the scene separated by $180°$. Weights were modified via an anti-Hebbian plasticity rule that weakens weights from active ring neurons onto active compass neurons (see Eq. 39 for the update rule). We updated these weights during saccades and in proportion to the squared velocity of each saccade, assuming that this velocity was constant through the duration of each saccade. In Fig 4c, and in the right column of Fig 4d, we used this weight matrix to determine the stability of the compass heading by allowing the compass bump to jump by $180°$ following a saccade; the probability of a jump was determined by comparing the net inhibition from active ring neurons at the orientation of the compass heading versus the symmetric (i.e., $180°$ shifted) orientation.

All weight matrices in Fig 4b-d were generated for a single simulation that evolved from a randomly-initialized weight matrix. In Fig 4d, we froze the final weight matrix and used it to generate many short simulations (mimicking the trial structure in the learning assay). We used these simulations to measure the average duration of fixations and turn bias of saccades as a function of heading. For the symmetric scene (right column of Fig 4d), we also measured the probability that the bump would jump from any given orientation in the EB. Fig 4e shows a snippet of the heading and arena trajectories from one such simulation.

Simulating the co-evolution of two learning systems. Fig 5 shows simulations in which both the compass weights and the goal weights are evolving over time. For all simulations, we initialized the compass weights in an asymmetric environment (the weight matrix in the left column of Fig 5b shows one such set of compass weights). For this initialization, we assumed that individual model flies maintained a fixed set of goal weights that was sinusoidal in form, with a strength that was sampled from one of 5 evenly spaced values between 0 and 1 (noninclusive), and with an orientation that was centered at either $\theta = 0$ or $\theta = \pm\pi$ (i.e., aligned with what would become the center of the danger zone). The goal activity in the left column of Fig 5b shows the readout from one such example of goal weights that have a strength of 0.67 and are centered at $\theta = 0$. After initializing the compass weights, we introduce model flies to an environment that mimics the environment used in the learning assay, in which model flies are punished whenever they orient toward one of two repeating patterns in a visual scene. Fig 5b shows the evolution of the compass and goal weights in this new environment for a single model fly; Fig 5c summarizes this evolution

over 1000 model flies. In both cases, we track the most stable compass heading (computed as the circular mean of $(1 - P_{jump}(\theta))$, and the circular mean of the goal activity. In Fig 5c, PI scores were computed by freezing the compass and goal weights at a given time, running a separate simulation with these frozen weights, and using this residency within this simulation to compute an estimate of the PI score.

The main panel of Fig 5d tracks the average strength of the goal heading, and the average angular distance between the goal heading and the center of the safe zone. Averages were performed over groups of 200 model flies that share the same initial goal strength. The lower panels of Fig 5d show the corresponding trajectories for the goal headings of individual model flies, colored by the angular difference between the goal heading and the most stable compass heading. Model flies were randomly selected based on their initial goal heading, and based on whether they had a high or low PI score at the end of training (upper and lower rows of model flies in Fig 5d, respectively).

Figs 5e-g track the strength of the goal heading, the alignment between the goal and most stable compass headings, and the PI scores over time for all model flies. In all cases, we temporally aligned these trials to the time of the weakest goal heading (vertical dashed lines in Figs 5e-g), and we sorted trials by the weakest goal heading. We then grouped these ordered trials into equally-sized groups of 200 model flies, and averaged the same quantities over trials within each group; these averages are shown in the lower panels of Figs 5e-g.