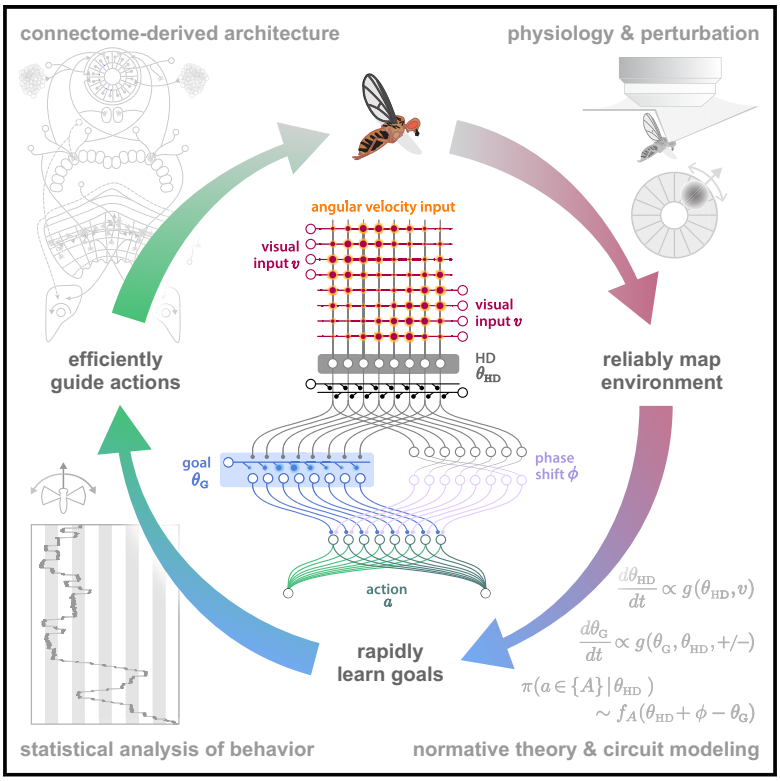
Neuron

Article

A neural circuit architecture for rapid learning in goal-directed navigation

Graphical abstract



Highlights

- Flies need head direction (HD) cells for operant goal learning in visual scenes with heat
- Behavior during learning is shaped by co-evolving internal HD and goal representations
- Structured but plastic neural circuits provide inductive biases for rapid learning
- The stability of internal representations shapes individual variability in learning

Authors

Chuntao Dan, Brad K. Hulse, Ramya Kappagantula, Vivek Jayaraman, Ann M. Hermundstad

Correspondence

vivek@janelia.hhmi.org (V.J.), hermundstada@janelia.hhmi.org (A.M.H.)

In brief

In novel environments, animals must simultaneously map their surroundings and form goals within them. Dan et al. combine anatomy, physiology, perturbation, and behavior to show how genetically specified circuit architectures with localized plasticity couple multiple evolving internal representations to make this learning fast yet flexible.

Dan et al., 2024, Neuron *112*, 2581–2599
August 7, 2024
https://doi.org/10.1016/j.neuron.2024.04.036

CellPress

Neuron

CellPress
OPEN ACCESS

Article

A neural circuit architecture for rapid learning in goal-directed navigation

Chuntao Dan,[1] Brad K. Hulse,[1] Ramya Kappagantula,[1] Vivek Jayaraman,[1,*] and Ann M. Hermundstad[1,2,*]
[1]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA
[2]Lead contact
*Correspondence: vivek@janelia.hhmi.org (V.J.), hermundstada@janelia.hhmi.org (A.M.H.)
https://doi.org/10.1016/j.neuron.2024.04.036

SUMMARY

Anchoring goals to spatial representations enables flexible navigation but is challenging in novel environments when both representations must be acquired simultaneously. We propose a framework for how *Drosophila* uses internal representations of head direction (HD) to build goal representations upon selective thermal reinforcement. We show that flies use stochastically generated fixations and directed saccades to express heading preferences in an operant visual learning paradigm and that HD neurons are required to modify these preferences based on reinforcement. We used a symmetric visual setting to expose how flies' HD and goal representations co-evolve and how the reliability of these interacting representations impacts behavior. Finally, we describe how rapid learning of new goal headings may rest on a behavioral policy whose parameters are flexible but whose form is genetically encoded in circuit architecture. Such evolutionarily structured architectures, which enable rapidly adaptive behavior driven by internal representations, may be relevant across species.

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 available, and many animals have evolved the ability to behave more flexibly by 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 is often conceptualized as operating in two phases: first, a latent learning phase in which an animal builds internal representations of its spatial relationship to the environment; and second, a phase in which the animal uses good or bad experiences to learn representations of goals anchored to these spatial representations and then modifies its behavior appropriately.[10] Many studies of learned behavior and its neural correlates, particularly in mammals, focus on the latter part of 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 Huber et al.,[11] Poort et al.,[12] Peters et al.,[13] Coddington and Dudman,[14] and Kuchibhotla et al.[15]). By contrast, in many natural settings, animals must develop spatial and goal representations simultaneously. Thus, the dynamics of one evolving representation impact other representations that are built upon it. Further, animals use these still-evolving representations to select appropriate actions, which in turn shape how these representations develop.

We delve into the processes by which animals simultaneously map a new environment and learn goals within it. We explore how these two learning processes interact to guide flexible behavior, how hardwired circuit motifs accelerate learning, and how the evolving dynamics of learning can shape individual variability in performance. We use a visually guided operant learning paradigm for the fly, *Drosophila melanogaster*,[16] to study how internal representations of HD and goals guide behavior (Figure 1A). In this paradigm, flies modify their actions in response to heat punishment associated with different instances of a repeating visual pattern. This visual symmetry alters the dynamics of flies' HD representation, allowing us to probe how its evolution affects and interacts with an internal goal representation to shape flies' behavior. By exploring how these processes might be implemented within an insect brain region called the central complex (CX)[17–24] (Figure 1B), we show how hardwired, modular circuit motifs provide strong inductive biases[25,26] for rapid learning in goal-driven behavior.

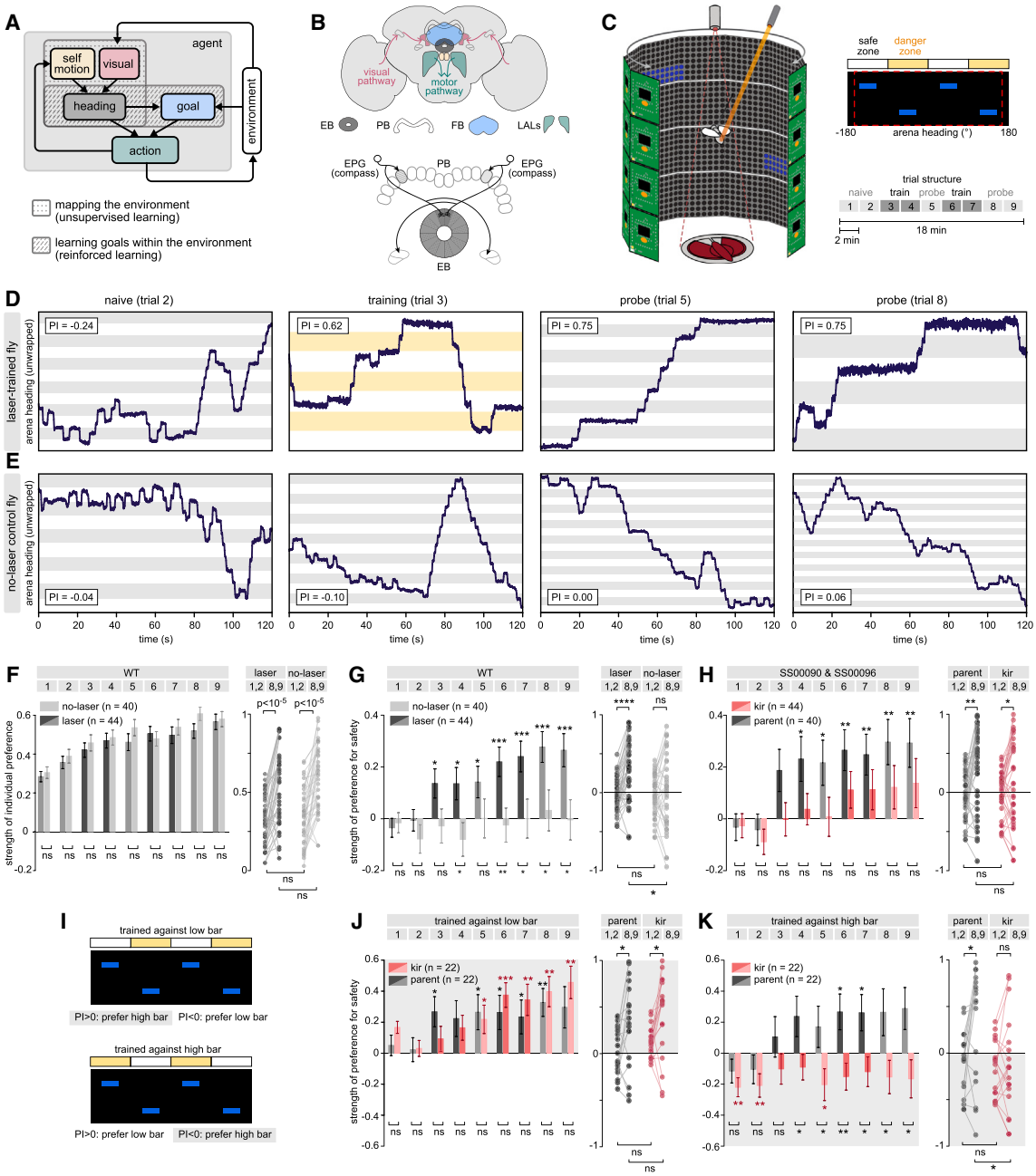


Figure 1. Operant visual learning requires an intact neural compass

(A) Two learning systems interact to guide behavior.

(B) The fly central complex (CX), highlighting the ellipsoid body (EB), protocerebral bridge (PB), fan-shaped body (FB), and lateral accessory lobes (LALs). Individual EPG (compass) neurons innervate a single "wedge" of the EB and a single "glomerulus" of PB; the population tiles the EB and PB.

(C) Left: flight simulator and LED arena used for behavioral experiments (method details). Upper right: during training, two symmetric and opposing quadrants are punished ("danger zone"; orange bars); the other two quadrants remain unpunished ("safe zone"; white bars). The red dashed box indicates the span of the visual arena (method details). Lower right: training protocol.

(D) Unwrapped heading trajectories and PI scores from a single laser-trained fly. Colored bands indicate arena headings that are punished (yellow bands) or that will be/have been punished (gray bands).

(E) Same as (D) for a no-laser control fly.

(F) Strength of individual preferences without regard to safety or danger (method details), measured for one genotype ("WT"). Left: fly-averaged preferences in all trials. Error bars: mean ± SEM. Significance: two-sided Wilcoxon rank-sum test. Right: individual preferences in early versus late trials (numbered boxes). Significance within groups: paired, two-sided Wilcoxon signed rank test. Significance between groups: two-sided Wilcoxon rank-sum test.

(G) Same as (F), but measuring strength of preferences for safety.

(legend continued on next page)

RESULTS

Tethered flying flies change their visually guided behavior after thermal conditioning

To study how flies adapt their behavior in new surroundings, we modified a well-established visual learning paradigm for tethered flying flies.[27] Flies were given closed-loop control of their angular orientation relative to a visual scene by using differences in their left and right wingbeat amplitude as a proxy for intended yaw movements (Figure 1C, left)[28,29] (method details). We used a periodic visual scene consisting of four quadrants, each containing a single horizontal bar (Figure 1C, upper right). 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 during a pair of 2-min-long "naive trials" (Figure 1C, bottom right). During subsequent "training trials," two symmetric quadrants (the "danger zone") were paired with an aversive laser punishment delivered to the abdomen of the fly; the remaining two quadrants (the "safe zone") were left unpunished. In "probe trials" with no punishment, we assessed whether flies formed lasting associations between different quadrants and the punishment.

Prior to training, flies explored different parts of the visual scene (Figures 1D and 1E, left columns). However, during training trials, flies typically avoided the danger zones (Figure 1D, middle column). In probe trials, flies continued to avoid the danger zones even after punishment was removed (Figure 1D, right two columns). In contrast, "no-laser" control flies that were not punished continued to explore different parts of the visual scene throughout all trials (Figure 1E), consistent with previous results.[16,27,30]

Previous studies have quantified such behavior using a performance index or "PI score"[31,32] that measures a preference for safety or danger. We first measured flies' heading preferences independent of safety or danger. We found that both laser-trained and no-laser control flies expressed a preferred arena heading even in naive trials, and they showed a strengthening of this preference across trials (Figures 1F and S1; note that we use the terms "heading" and "HD" interchangeably in our head-fixed setting). This strengthening might result from flies improving the control of their saccades and fixations in this new visuomotor setting; in free flight, visual textures are known to influence the structure of flies' behavior.[33,34] Notably though, only laser-trained flies significantly shifted their preferred arena headings toward safety (Figure 1G, left column, dark bars), consistent with sample trajectories (Figures 1C and 1D) and with past results.[31,32,35] No-laser control flies did not significantly change their preference for either quadrant (Figure 1G, left column, light bars). These preferences were also reflected at the level of individual flies (Figure 1G, right column).

We next sought to test whether flies' ability to flexibly modify their heading preferences depends on an intact representation of heading. Past studies have shown that perturbing various CX neuron types significantly impacts flies' performance on this task.[35] The HD representation itself is maintained in the dynamics of "EPG" or "compass" neurons[36] in a CX structure called the ellipsoid body (EB). Compass neurons[37–39] 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; these inputs have been also linked to flies' ability to remember specific orientations relative to a disappearing visual landmark.[41,42] We thus asked whether compass neurons are required for this operant learning. To test this, we silenced compass neuron activity by selectively expressing the inwardly rectifying potassium channel Kir2.1 in compass neurons, using two different split-GAL4 lines (Figures 1H–1K, S2, and S3). These flies and flies from their parental control groups did not fly as well as wild-type (WT) flies (data not shown), but they displayed normal heading preferences in single-stripe environments (Figure S4; method details). Control flies showed high PI scores in training and probe trials (Figure 1H, left, gray bars). In contrast, flies with silenced compass neurons consistently showed low PI scores (Figure 1H, left, red bars). However, these average trends were not entirely reflected in the behavioral preferences of individual flies; both parental control flies and compass-neuron-silenced flies showed significant shifts in their PI scores after training (Figure 1H, right).

To reconcile these trends, we asked whether compass-neuron-silenced flies might be expressing an innate preference for a sun-like stimulus. Previous results from both walking and flying flies have shown that silencing compass neurons or visual inputs to the compass neurons exposes hardwired phototactic behaviors[37–40] that are likely sensitive to the specific shapes and positions of visual stimuli.[43] In particular, flying flies are known to fly directly toward a sun-like stimulus when their compass neurons are silenced.[37] We thus asked whether the high bar's resemblance to a sun-like stimulus, together with a general increase in the preference for this stimulus (Figure 1F), might account for our unexpected results. An innate preference for high bars would manifest in positive PI scores when trained against low bars and negative PI scores when trained against high bars (Figure 1I). This is indeed what we observed when we silenced compass neurons (Figures 1J and 1K): flies exhibited a consistent preference for the high bar, regardless of which pattern they were trained against. In contrast, parental control flies showed consistent preferences for safety. Thus, when compass neurons are silenced, it is likely that flies' behavior in this paradigm is driven by visuomotor pathways that do not involve the CX. Taken together, these experiments suggest that WT flies quickly learn to avoid dangerous parts of the visual scene in this paradigm and that an intact HD representation is required for normal visual learning.

(H) Same as (G), but for double-blind experiments in which two different genotypes of flies ("SS00090" and "SS00096") underwent laser training (method details). One group ("Kir") had Kir expressed in EPGs; a second group of genetically matched flies did not ("parent"). See Figures S2B–S2F for PI scores split out by genotype.

(I) We split training conditions by whether the low versus high bar was punished.

(J and K) Same as (H), but split by training conditions. $^{*}p \leq 0.05$; $^{**}p \leq 0.01$; $^{***}p \leq 0.001$; $^{****}p \leq 0.0001$.

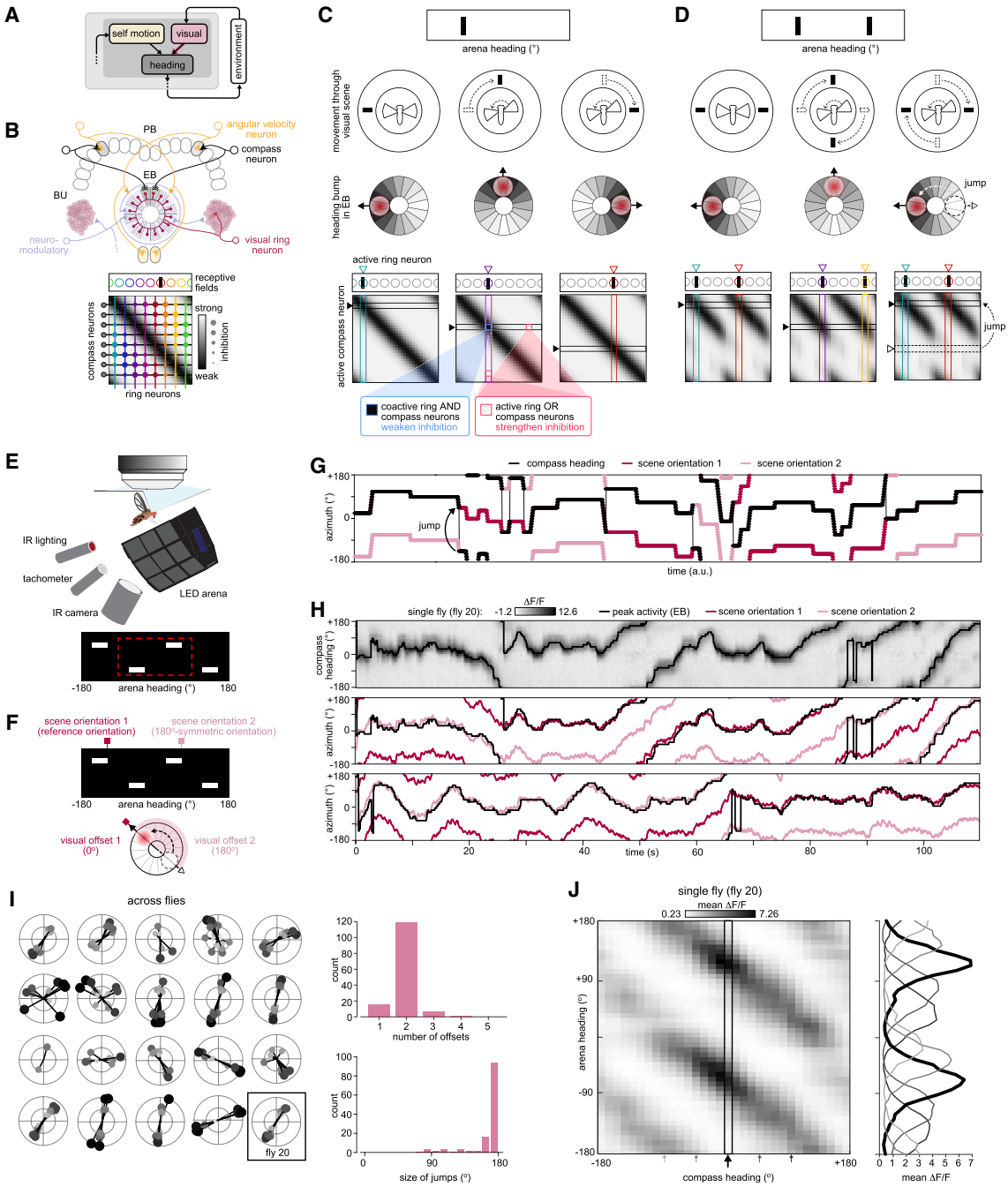


Figure 2. Symmetric scenes trigger jumps in compass neuron dynamics

(A) Unsupervised learning tethers the compass heading to the visual environment.

(B) Upper: neuromodulatory neurons mediate plasticity in all-to-all inhibitory synapses (filled red circled) from visual ring neurons onto compass neurons. Lower: a self-consistent mapping is marked by diagonal structure in the weakened inhibition from ring neurons onto compass neurons.

(C) In an asymmetric scene with a single visual pattern, plasticity stabilizes a self-consistent mapping between active ring neurons and active compass neurons. When the fly makes a saccade (different columns), the heading bump remains tethered to movements of the scene.

(D) Same as (C), but for a symmetric scene with two repeating visual patterns separated by 180°. Plasticity stabilizes a 2-fold symmetric mapping (repeating diagonal bands in the weight matrix). When the fly makes a saccade (different columns), the compass bump can jump to an EB location that is only that weakly, rather than strongly, inhibited by active ring neurons.

(E) Two-photon calcium imaging setup.

(F) For a 180°-symmetric scene, the bump could maintain two "visual offsets" relative to a reference scene orientation; these would be separated by 180° and match identical views of the scene.

(legend continued on next page)

Visual symmetries trigger jumps in flies' HD representation

Having established that learning in this paradigm depends on an intact HD representation, we next sought to understand how this representation co-evolves with behavior (Figure 2A). We consider two separate but interacting learning processes: we first study how unsupervised learning governs changes in the HD representation as it tethers to a visual scene with symmetries (Figure 2); we then study how reinforcement learning (RL) guides changes in behavior during thermal conditioning (Figures 3 and 4); finally, we study how these two learning systems interact to shape individual variability in performance (Figures 5 and 6).

The symmetry of our visual setting, a simplification of the symmetries observed in natural scenes,[44] creates a unique opportunity to study how an internal HD representation develops over time and impacts the learning of goal-directed behavior. The fly's HD representation is maintained as a single localized bump of activity that typically rotates around the EB in concert with the fly's rotations through a visual scene.[36] This self-consistency arises via plasticity between the compass neurons that maintain the bump in the EB and so-called ring neurons that bring visual inputs into the EB. Visual ring neurons have feature-tuned receptive fields that tile space,[45–47] and they synapse onto compass neurons via all-to-all inhibitory connections in the EB[24] (Figure 2B). During exploration of a new visual scene, inhibitory Hebbian-like plasticity weakens synapses from active ring neurons onto active compass neurons at the location of the compass bump in the EB,[44,48] and it is modulated by the fly's angular velocity[44] via dopaminergic neurons that innervate the EB.[24,49–51] Over time, this plasticity acts in an unsupervised manner to create a self-consistent mapping between the visual scene and the HD representation; in a simple scene with a single landmark, this self-consistency would be characterized by a diagonal band in the matrix of synaptic weights between ring and compass neurons (heatmaps in Figures 2B and 2C).[44,48] Once stabilized, this mapping would ensure that the bump moves in synchrony with the visual scene, such that the fly's rotations through the scene activate ring neurons with the appropriate spatial receptive fields (Figure 2C).

To study how these dynamics change in symmetric scenes, we built a circuit model that captures these key ingredients (Figures 2C and 2D; method details). In a simple scene with 2-fold symmetry, the plasticity described above creates a mapping with two bands, such that ring neurons with 180°-opposite receptive fields have approximately equal synaptic weights onto the same compass neurons (Figure 2D, heatmaps). Since the scene is symmetric, these ring neurons will be identically active at two arena headings separated by 180°. However, the corresponding compass neurons—whose HD tuning is separated by 180°—will be inhibited to different degrees. This can trigger a competition between two sets of compass neurons that are activated by the same ring neurons,[36,44,48,52] and the bump can jump across the EB to the location that is most weakly inhibited by the same active ring neurons (Figure 2D, right column).

Such jumps are not restricted to visual scenes with precise symmetries[44]; a scene with similar visual features at multiple locations would likewise evoke similar ring neuron activation patterns at multiple arena headings, which in turn would trigger a competition between multiple compass neurons that are tuned to those headings. In the specific case of a 2-fold symmetric scene, we assume that the bump jumps between two locations separated by 180°, with a probability determined by the difference in net inhibition between the two locations.

To explore whether the HD representation in real flies exhibits similar dynamics, we used two-photon calcium imaging to monitor the HD representation in tethered flying flies in a visual setting, similar to that used in the learning assay (Figure 2E; method details). We measured the "visual offset"[36,38,44,48,52] between the compass bump and symmetric orientations of the visual scene (Figure 2F); this offset varies over time and from fly to fly.[36–38,44,48] Consistent with our circuit model (Figure 2G), we found that the compass bump tended to jump between two offsets that reflected symmetric views of the visual scene (Figure 2H). The distribution of offsets was bimodal in a majority of flies (Figure 2I, left), with two peaks separated by 180° (Figure 2I, right). In correspondence with this, different wedges of the EB were active for symmetric views of the visual scene, and thus their heading tuning curves had two peaks separated by 180° (Figure 2J). This resulted in a two-to-one mapping from the visual scene onto the HD representation, similar to our model (Figure 2D) and to the tuning previously observed in simpler symmetric scenes.[36,44,48,52] Thus, in the 2-fold symmetric scenes used in our learning paradigm and in previous experiments,[31,32,35] the HD representation likely jumps between orientations that correspond to symmetric views of the scene.

A probabilistic policy captures tethered flies' visually guided behavior

Having established how the fly's HD representation behaves in the visual setting of our paradigm, we next asked how this representation is used to guide learned changes in behavior. We sought to construct a generative model of behavior, or behavioral "policy,"[53] that could capture naive and conditioned behavior. During free and tethered flight, flies exhibit periods of fixation during which they maintain a near-constant heading over time,[28,54,55] punctuated by ballistic turns, or body "saccades," that drive abrupt changes in heading[54,55] (Figure 3A). We approximated behavior as being composed of only these

(G) As a model fly navigates with respect to the scene in (D), the compass heading jumps between symmetric views of the scene.

(H) First row: compass neuron calcium activity during closed-loop tethered flight with a symmetric visual scene. Second and third rows (trials 1–2): the peak activity jumps between two offsets that correspond to symmetric views of the scene.

(I) Left: offsets for all trials for each fly. Orientations indicate offset value; marker size, color, and radius indicate fraction of time spent at each offset (outer radius equals 1). Right: number of unique offsets (top) and angular size of jumps (bottom), aggregated across flies and trials. A majority of jumps match the 180°-symmetry of the visual scene; see additional resources for discussion of the smaller peak at 90°.

(J) Main: average $\Delta F/F$ of compass neuron activity in different wedges in the EB (compass headings) as a function of arena heading, averaged across all nine trials for the fly shown in (H). Right: tuning curves for compass neuron activity in different EB wedges (denoted by arrows in the main panel).

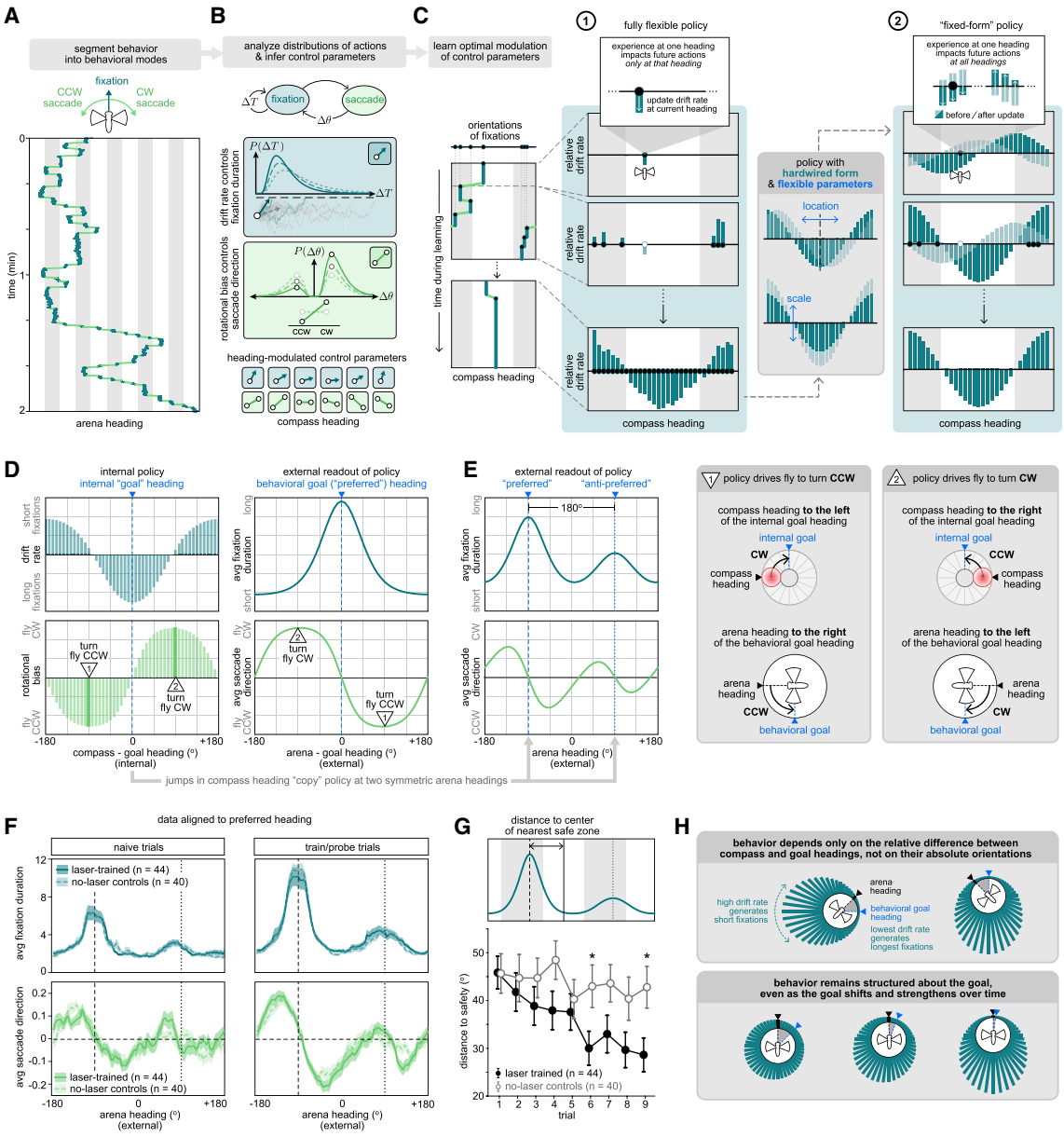


Figure 3. An inferred policy captures conditioned and unconditioned behavior

(A) Left: behavioral trace from Figure 1D, segmented into fixations and saccades.

(B) A behavioral policy generates fixations of varying duration ΔT and saccades of varying size $\Delta\theta$. Shaded boxes: $P(\Delta T)$ is well fit by a drift-diffusion process with an adaptive drift rate (arrow icon); $P(\Delta\theta)$ is well fit by a lognormal distribution with an adaptive rotational bias (barbell icon). Distributions are schematized; see Figure S5 for data distributions. Lower: we hypothesize that the fly's internal compass heading modulates the drift rate and rotational bias.

(C) Schematics of two behavioral policies whose control parameters could be modified over time to guide the behavior of an agent, or model fly (illustrated for the fixational drift rate). (C1) With a fully flexible policy, the agent must iteratively sample all headings to learn to generate low drift (long fixations) in the safe zone and high drift (short fixations) in the danger zone. (C2) The agent could instead use a behavioral policy whose form is fixed and resembles the final outcome of scenario (C1). Experiences at one heading would be used to modify the location and scale of this policy, and thus an association made at one heading would impact future actions taken at all other headings.

(D) Left: final behavioral policy of a reinforcement learning agent trained to maintain a single internal goal heading, using the policy schematized in scenario (C1) (method details). Right: external readout of internal policy. Numbers mark scenarios schematized in gray boxes to far right.

(E) When the policy in (D) is coupled to an internal compass heading, any jumps in the compass heading serve to "copy" the policy at symmetric arena headings. This leads to a bimodal behavioral readout.

(legend continued on next page)

two modes (Figures 3A and 3B) and determined transitions between them (Figure S5A; method details). Individual fixations and saccades vary in their angular velocity and duration (Figure S5B); we used this variability to infer a generative model in which flies control higher-dimensional distributions of actions through a lower-dimensional set of control parameters. Specifically, our analysis supports a generative model in which flies control the average direction of saccades through an adaptive rotational bias and the average duration of fixations through a drift-diffusion process with an adaptive drift rate (Figures 3B middle rows and S5C–S5J). The rotational bias and drift rate thus act as control parameters that flies can use to favor turns in one direction over another and to fixate more or less. We hypothesized that learning could modify these control parameters based on the fly's internal compass heading (Figure 3B, bottom row).

How should these parameters be modified over time to appropriately control fixations and saccades? Figure 3C schematizes two distinct policies that could be modified through experience to generate lower drift rates, and thus longer fixations, in the safe zone. One policy stores the value of different parameter settings at each compass heading, similar to an RL algorithm called Q-learning.[53] In its simplest formulation, an agent (i.e., a model fly) associates individual headings with individual action parameters, and it independently updates these associations based on reinforcement. The resulting policy can flexibly take on a variety of functional forms. However, this flexibility comes at the cost of being slow to learn, because the agent needs to iteratively sample all headings to learn a complete set of associations. In our setting, this policy converges to a profile of drift rates that is lowest—and thus generates the longest fixations—in the center of the safe zone (Figure 3C1). Before and during learning, this policy has no guarantee of selecting appropriate actions for headings that the agent did not directly sample.

Alternatively, more rapid learning might be enabled by *fixing*, rather than *learning*, the functional form of the policy. The ideal form of such a policy might be expected to resemble the form to which a fully flexible policy converges after training (Figure 3C1, bottom row). Learning would then act to shift and scale this function, rather than build it up from individual associations. As a result, reinforcement experienced at one heading would immediately impact the selection of actions at all other headings. This type of non-local policy update would accelerate the learning process, but it would limit the complexity of the heading-dependent associations that can be learned. The control parameters would have similar profiles—and therefore generate similar distributions of actions—before, during, and after learning, but with experience-dependent shifts in their location and scale (Figure 3C2).

We reasoned that such a fixed-form policy, if tethered to a preferred "goal" heading, might account for flies' behavior in our paradigm. Individual flies, like many other insects, maintain a preferred heading for periods of time even when tethered (Figure S1),[37,38,40] and they locally explore headings centered around their preferred heading.[37,38,40,56–58] This "menotaxis" behavior is thought to aid dispersal and long-range navigation.[59–64] Consistent with these results, when we trained model flies to maintain a single goal heading, they learned to minimize the drift rate of fixations at the goal heading and to bias their saccades toward the goal heading (Figure 3D; additional resources). We hypothesized that the resulting behavioral patterns could constitute a fixed-form policy that depends only on the *relative difference* between the fly's compass and goal headings. Because the symmetric scene in our paradigm leads to jumps in the compass heading (Figure 2), we would expect the form of this policy to be "copied over" at two symmetric arena headings (Figure 3E; method details). With such a policy, a fly's internal goal heading would determine its probabilistic behavior at *all* arena headings, regardless of laser training. Moreover, if the compass bump jumps by 180°, the fly would behave similarly at opposite headings.

Consistent with these expectations, when we aligned the fixations and saccades from individual flies to their preferred arena headings—a proxy for their internal goal headings—we observed bimodal behavioral curves, with flies locally directing their saccades toward and fixating longer at the two orientations that correspond to symmetric views of the visual scene (Figure 3F). As expected, both naive and laser-trained flies exhibited this same behavioral structure about their preferred headings, but only laser-trained flies reliably shifted these preferences toward the center of the safe zone (Figures 3G and S6).

Together, these results suggest that flies rely on a behavioral policy whose form is fixed and whose flexibility arises from internal representations of compass and goal headings that are shaped by the visual environment and by punishment. For this policy to retain its fixed form, circuits in the fly brain must ensure the following: (1) that the fly's actions depend only on the relative difference between compass and goal headings, and not on their absolute orientations (Figure 3H, left); and (2) that the behavioral structure about the goal heading remains intact even as the goal heading is changing over time (Figure 3H, right).

A connectome-inspired circuit model can structure a goal-driven behavioral policy

To understand how the fly's circuitry might enforce a fixed-form behavioral policy (Figure 3H), we constructed a second circuit model that combines two internal representations—of compass and goal headings—to guide the fly's actions (Figure 4A). In this model, the goal heading evolves over time via a reinforced learning process, and it is compared with the compass heading to drive behavior. This model combines insights from existing models[65–68] and from the fly CX connectome,[24] but it intentionally abstracts away details of CX circuitry to focus on key computations that we believe underlie flies' behavior in this assay.

(F) Average fixation duration (upper) and saccade direction (lower) generated by laser-trained and no-laser control flies. All data are aligned to the preferred arena headings of individual flies (vertical dashed line). Shaded regions: mean ± SEM. See Figure S6 for data aligned to safety/danger.

(G) Distance between preferred arena headings and the center of the safe zone. Significance: two-sided Wilcoxon rank-sum test (*$p \leq 0.05$). Error bars: mean ± SEM.

(H) Minimal algorithmic requirements to tether a fixed-form behavioral policy to a flexible goal heading.

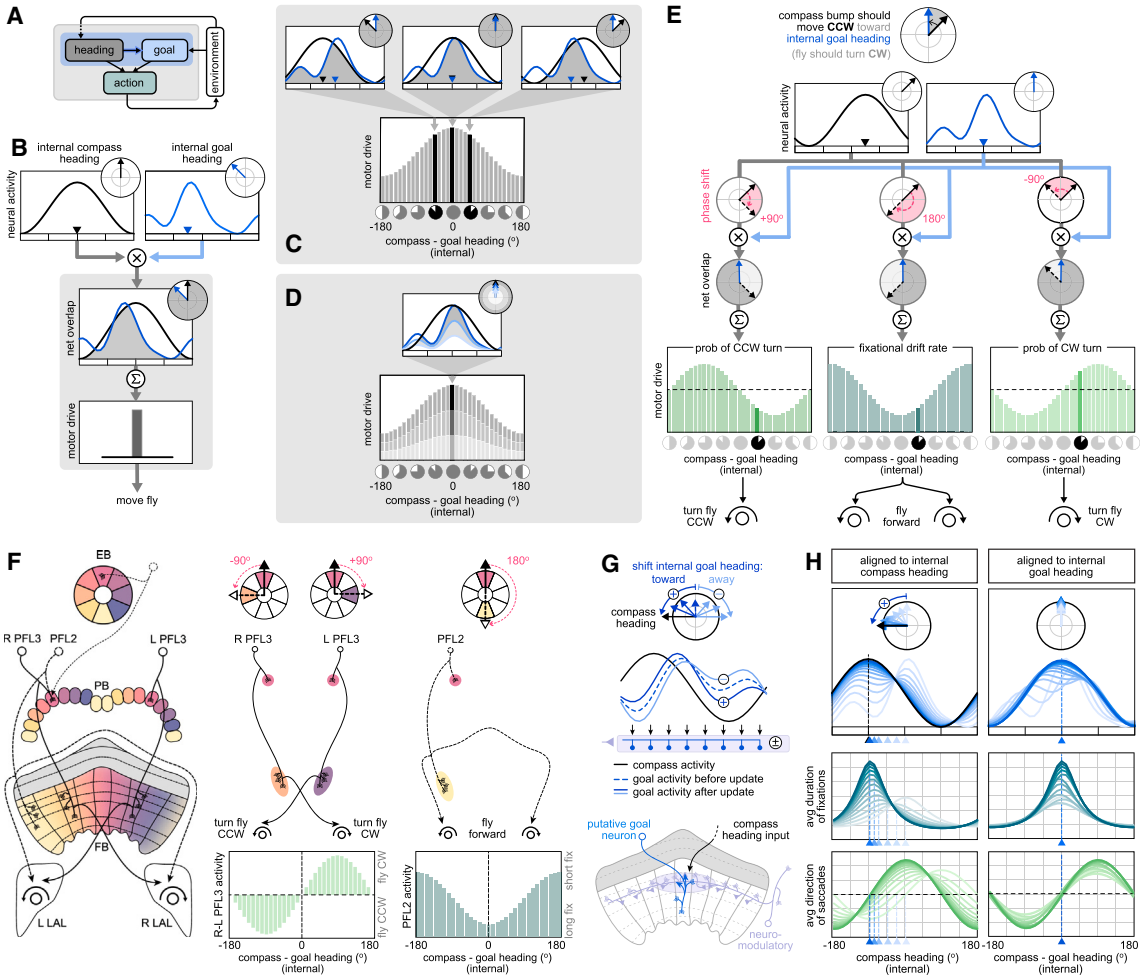


Figure 4. Hypothesized implementation of a fixed-form behavioral policy tethered to a flexible goal heading

(A) Reinforcement learning updates a flexible goal heading to drive behavior.

(B) We multiply a sinusoidal compass activity profile and an arbitrarily shaped goal activity profile; the summed output (similar to the net overlap between profiles; gray) is used to drive rotational controllers. Phasors denote circular means of compass and goal profiles.

(C) Different orientations of a compass heading (black curve) relative to a fixed goal (blue curve) generate different motor drives (filled bars). Sweeping the compass heading yields a sinusoidal motor drive that peaks when the compass heading is aligned with the goal heading.

(D) Weakening the goal profile reduces the motor drive across all compass headings.

(E) Phase-shifting the compass heading by $+90°$, $+180°$, and $-90°$ results in motor drives whose peaks are shifted relative to the goal. If these are used to drive rotational controllers, we recover the fixed-form policy schematized in Figures 3D and 3H.

(F) Left: PFL neurons could implement the phase shifts and premotor projections schematized in (E). These are a simplification of true phases; see Hulse et al.[24] Neurons in each PFL population display similar motifs and tile the FB, covering all angles. Middle: two populations of PFL3 neurons implement $\pm$ 90° phase shifts and project unilaterally to one of two premotor regions (the LALs); the difference in activity of these two populations could control the probability of CW or CCW saccades. Right: the PFL2 neuron population implements a 180° phase shift and projects bilaterally to both LALs; the activity of this population could control the drift rate of fixations.

(G) Neuromodulatory neurons could modify the goal heading based on the fly's current compass heading. Positive reinforcement and negative reinforcement shift the goal heading toward or away from the compass heading, respectively.

(H) Left column: evolution of the goal profile (blue curves) when a model fly experiences positive reinforcement at a fixed compass heading (black arrows/curves). Over time, the goal profile is strengthened at the current compass heading and weakened away from it, thereby shifting the goal heading. The behavioral readouts (green curves) shift with the goal. Right column: same as left, but aligned to the goal.

In Figure S7 and additional resources, we link individual model components to potential CX neurons and circuit motifs.

To match our behavioral observations (Figure 3F), a key requirement of our circuit model is that it produces drift rates and turn biases that vary approximately sinusoidally with the relative difference between the compass and goal headings (Figure 3D). This sinusoidal structure can be achieved by computing the dot product between two circular activity profiles: a sinusoidal compass activity profile and an arbitrarily shaped goal activity profile (Figure 4B). When repeated for all possible orientations of the goal activity profile, the output is itself sinusoidal, regardless of the specific profile of goal activity (Figure 4C).

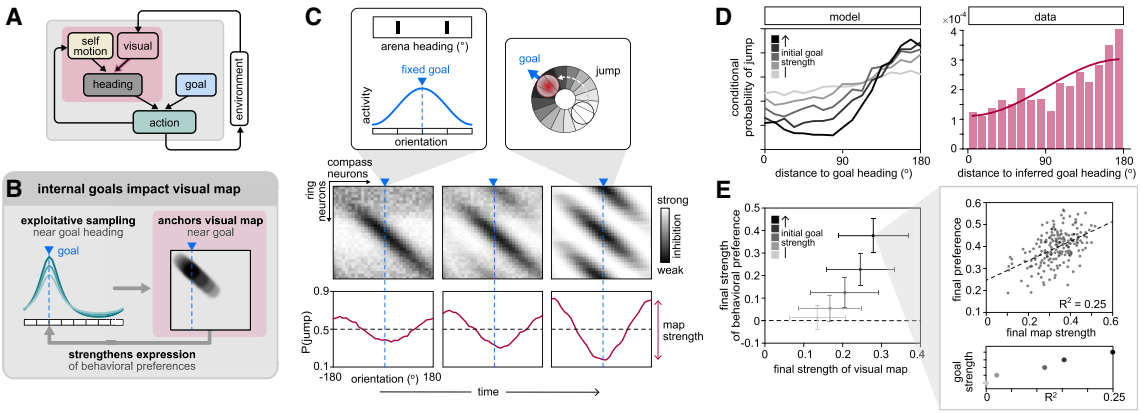


Figure 5. The internal goal shapes the construction of a visual map

(A) The goal impacts the visual map through the actions selected by the behavioral policy.

(B) The behavioral policy leads to exploitative sampling around the goal heading, which impacts the evolution of the visual map and, in turn, the strength of behavioral preferences around the goal.

(C) Evolution of a visual map for a single model fly that uses a fixed goal heading to navigate in a symmetric visual scene. The map was initialized in an asymmetric scene. Over time, the map develops repeating bands of weakened inhibition, and the most stable compass heading—at which there is the lowest probability of a bump jump—aligns with the goal heading.

(D) Left: stronger initial goal headings lead to more pronounced differences in the conditional probability that the bump will jump when it is close to versus far from the goal heading. Right: in real flies, the conditional probability of a bump jump increases as a function of angular distance from the preferred compass heading (method details). Bars: average probability across flies. Solid line: best-fitting cosine curve.

(E) Stronger initial goal headings lead to stronger visual maps, which strengthen behavioral preferences about the goal heading. Error bars: mean ± SD.

Mathematically, this corresponds to taking the convolution between the compass and goal profiles; the output can then be used as a motor drive to modulate downstream rotational controllers. Under this operation, the circular mean of the goal profile specifies the goal heading at which the motor drive peaks (additional resources). Strengthening or weakening the goal profile changes the amplitude of the motor drive, but it does not alter its shape (Figure 4D). To shift this motor drive so that it peaks away from the goal heading, as required (Figure 3D), one can introduce a phase shift to the compass profile before combining it with the goal. Shifts by 90° produce the largest output when the compass heading is 90° to the right or left of the goal heading, and they are thus suitable to bias clockwise (CW) or counterclockwise (CCW) saccades toward the goal. Similarly, a shift by 180° has the largest output when the compass and goal headings are anti-aligned, and it is thus suitable to specify high drift rates that drive short fixations. Thus, this architecture can generate a fixed-form policy whose motor drives vary sinusoidally as a function of the difference between the compass and goal headings, as required.

This architecture combines three key ingredients: a sinusoidal compass activity profile, phase shifts of − 90°, + 90°, and + 180° applied to the compass activity profile, and an arbitrarily shaped goal activity profile that can be modified based on experience. The sinusoidal profile of the compass activity is ideally suited for performing vector computations[69] and is thought to be enforced by neurons in the protocerebral bridge (PB).[24,70] Phase shifts to the compass profile have been described by us and others[24,67,71] and have found support in recent physiological data.[72,73] These shifts are determined by propagating a potential activity bump from the compass neurons to their downstream partners in the PB and fan-shaped body (FB).[24,70] Neurons that receive input from similarly tuned neurons in the PB and that project to similar columnar regions of the FB are considered to have a 0° phase shift. These neurons serve as a reference frame for analyzing projection patterns of neurons that implement nonzero phase shifts. Based on such projection patterns, populations of "PFL" neurons are best suited to control fixations and saccades based on the fly's current heading. All three populations inherit a copy of the heading bump that is phase-shifted in the FB relative to the PB, and they project to premotor regions called the lateral accessory lobes (LALs), where they are well-positioned to control CW or CCW rotations.[24,67,71,74] The left and right PFL3 populations implement phase shifts of − 90° and + 90° project unilaterally to each LAL; these populations are thus ideally suited to drive CW and CCW saccades (Figure 4F, middle column; see also Pires et al.[72] and Westeinde et al.[73]). The PFL2 population implements a phase shift of 180° and projects bilaterally to both LALs; this population is thus ideally suited to drive fixations (Figure 4F, right column; Hulse et al.[24]; see also Matheson et al.[71] and Westeinde et al.[73]). Our circuit model assumes that the responses of all three populations are tuned to the difference between the compass and goal headings, as described above (Figure 3D); the difference between the population activity of the left and right PFL3 neurons determines the probability of a CW or CCW saccade, while the population activity of the PFL2 neurons determines the drift rate of fixations (Figure 4F, bottom row). The PFL2 and PFL3 populations mutually inhibit each other to prevent saccades during a fixation and vice versa.

Finally, we hypothesize that the goal profile is stored in synapses from a tangential neuron, which innervates a layer of the FB, onto putative columnar goal neurons. Tangential neuromodulatory neurons could then update the strength of these synapses

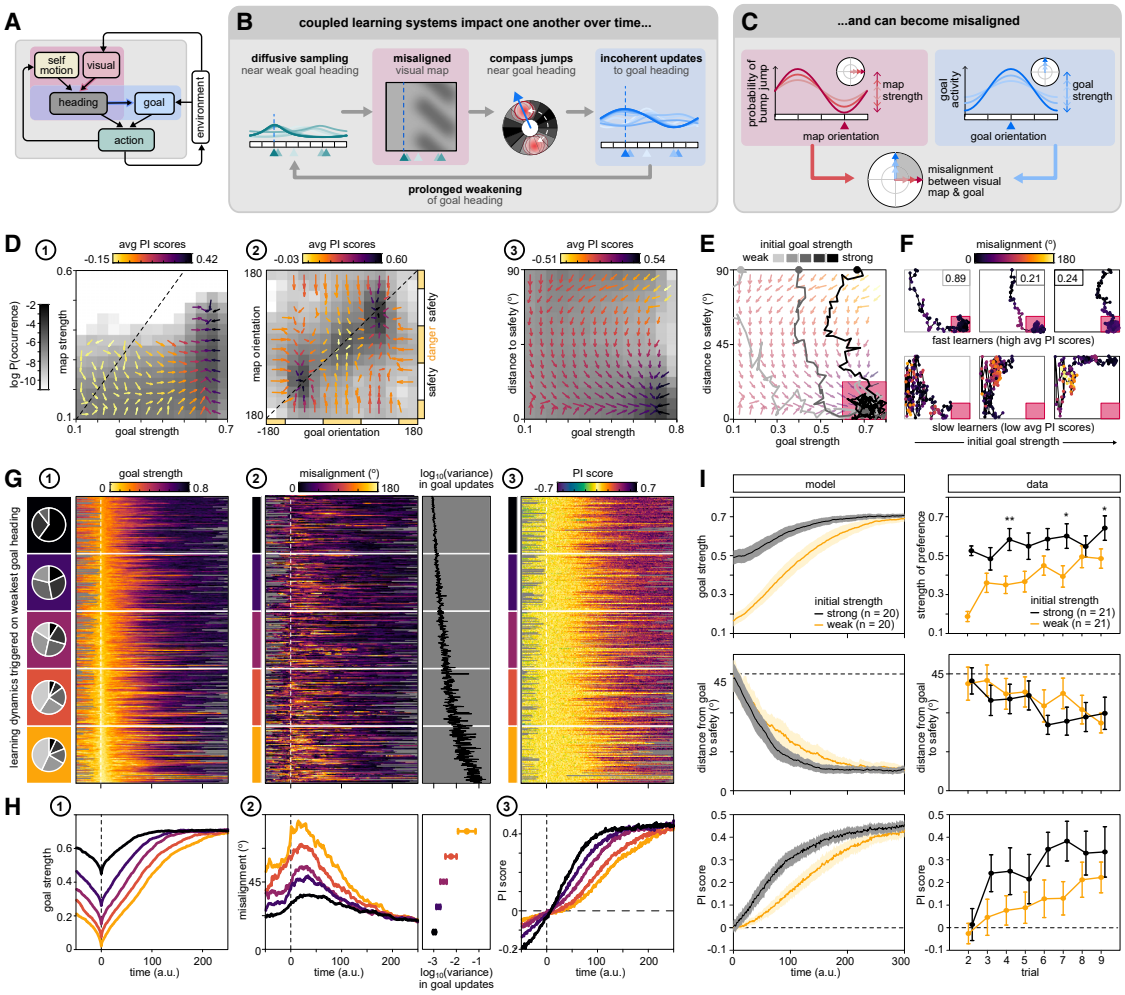


Figure 6. Visual scene mapping and goal learning interact to influence behavior

(A) Two learning systems (red and blue boxes) are coupled through the compass heading and the actions driven by the behavioral policy.

(B) A weak goal heading leads to diffusive behavior that can fail to anchor the visual map. This can cause the bump to jump near the goal heading, which in turn leads to incoherent updates of the goal heading, for example, if two 180°-opposite headings are successively reinforced or punished.

(C) We measure the evolving alignment between the visual map and goal heading.

(D) Vector maps illustrating the average effect of learning. Individual vectors show changes in the strengths and orientations of the visual map and goal heading, centered on their initial values; longer arrows denote stronger average effects of learning (method details). Vector flows illustrate the progression of learning.

(E) Trajectories of individual model flies, superimposed over the vector map from (D3). Red-shaded box indicates strong goals aligned with the safe zone.

(F) Same as (E), but trajectories are colored by the misalignment between the most stable compass heading and the goal heading (see C). Boxed numbers indicate the fraction of simulation time required to stably reach the red-shaded box; flies in the lower row did not reach the red box.

(G) Learning trajectories aligned by the time point when the goal heading was weakest (vertical dashed lines) and sorted by the value of the weakest goal heading. Colored rectangles indicate the groupings that were used to construct the averages in (H); pie charts indicate the fraction of model flies within each group that began with different initial goal strengths. Weak goals (G1) lead to misalignment between the most stable compass heading and the goal heading (G2 left), higher variance in the updates of the goal heading toward its final value (G2 right; also see Figure S9), and slower changes in PI scores (G3).

(H) Same as (G), but averaged over groups of model flies that exhibited similar dynamics in their goal headings. Error bars: mean ± SD.

(I) Left column: behavior of model flies that began with weak versus strong initial goal headings. Shaded regions: mean ± SD, computed over 100 sets of randomly selected groups of 20 model flies. Right column: same as left column, but for laser-trained WT flies that began with weak versus strong initial preferences. See Figure S10 for data combined across genotypes, where differences are more pronounced. Error bars: mean ± SEM. Significance: two-sided Wilcoxon rank-sum test (*$p \leq 0.05$; **$p \leq 0.01$).

via bi-directional Hebbian plasticity that is modulated by the fly's current compass heading (Figure 4G), an idea with anatomical support from the connectome.[24] Negative reinforcement would weaken goal synapses at the current compass heading and thereby push the goal heading away from it, whereas positive reinforcement would strengthen synapses at the current compass heading and pull the goal heading toward it (Figure 4H). The strengths of these synapses could then be read out by other columnar goal neurons before being transmitted to PFL neurons, which would combine this goal input with the phase-shifted

compass heading to drive behavior. As a result, the duration of fixations and directionality of saccades would remain structured about the goal heading, even as it evolves over time (Figure 4H, left column). As the goal heading strengthens, the fly's behavioral expression of this goal would also grow stronger, with longer fixations near the goal heading and more biased turns toward it (Figure 4H, right column). This would naturally lead to more exploratory, diffusive behavior when the goal heading is weak and more exploitative, goal-directed behavior when the goal heading is strong.

A fly's internal goal influences the mapping of visual scenes onto its compass

In previous sections, we showed how the symmetric visual scene of our paradigm is mapped onto an internal compass heading in the EB (Figure 2), how circuits in the PB and FB might implement a fixed-form behavioral policy tethered to the difference between the compass and goal headings (Figures 4E–4G), and how reinforcement signals in the FB might modify goal headings and thereby shift and scale the policy (Figures 4G and 4H). Next, we combined our two circuit models (Figures 2 and 4) to explore how actions selected by this policy impact the mapping of the visual scene onto the compass in the EB (Figure 5A).

As described earlier (Figure 2), the visual mapping from ring neurons onto compass neurons is influenced by repeating patterns in the visual scene: in our 2-fold symmetric setting, this mapping develops two bands, such that ring neurons with 180°-opposite receptive fields have approximately equal synaptic weights onto the same compass neurons. Importantly, when this mapping is driven by behavior that is tethered to a fixed internal goal heading, these patterns of synaptic weights develop at specific locations relative to that goal (Figure 5B). Specifically, because most of the fly's saccades are initiated in the neighborhood of the goal heading, compass neurons that are tuned to headings near the goal will be the first to develop weakened synapses from active ring neurons. This will create a difference in net inhibition between these compass neurons and those with 180°-symmetric tuning, which will cause the HD bump to favor regions of the EB near the goal heading. Whenever the bump is away from the goal heading, it will tend to jump back toward it, further weakening the synapses from active ring neurons near the goal. Over time, this positive feedback loop will ensure that the bump is least likely to jump when the fly's compass heading is aligned with the goal (Figure 5C). The stronger the goal heading, the more pronounced this effect (Figure 5D, left).

To examine whether real flies exhibit similar compass dynamics, we analyzed the location of bump jumps relative to an inferred goal heading, which we defined as the location of maximal residency in the EB that corresponded to the fly's preferred arena heading measured from behavior (method details). We found that jumps were least likely to occur at this inferred goal heading and most likely to occur 180° away from it (Figure 5D, right). Thus, symmetries of the visual scene induce jumps in the compass bump[36,44,48,52]; in a scene with 2-fold symmetry, these jumps are not uniform around the EB but are most likely to occur at the location symmetric to the goal heading in the EB (Figure 5D, right), as predicted (Figure 5D, left). Finally, when we analyzed the behavioral preferences of model flies, we found that stronger internal goals led to stronger visual maps, which in turn led to stronger behavioral preferences (Figures 5E and S8). This effect may contribute to the non-specific strengthening of preferences we observed in both laser-trained and no-laser control flies (Figure 1F).

Deciphering fly-to-fly variability in operant visual learning

Until this point, we have considered the separate learning processes of modifying an HD representation and modifying a goal representation. However, in our operant visual learning paradigm, the fly must form these two representations simultaneously (Figure 6A). These representations evolve over time via separate, but coupled, learning processes: the mapping of the visual scene onto compass neurons evolves via an unsupervised learning process modulated by the fly's saccades through the visual scene, while the goal heading evolves via a reinforced learning process guided by positive and negative experiences at different compass headings. These learning processes are coupled through the compass heading, which directly updates the goal heading, and through the behavioral policy, which is tethered to both the compass and goal headings and which determines the visual and thermal feedback via the actions that the fly selects (Figure 6B).

To explore how behavior might be influenced by the interactions between these two learning processes, we combined our two circuit models in a simulated task that mimicked what real flies experienced, with a 2-fold symmetric scene coupled to negative and positive reinforcement. We then tracked the orientation and strength of both the visual map and the goal heading over the course of learning (Figure 6C). Prior to training, each model fly began with a goal heading that was used to form a visual map in an asymmetric scene; for simplicity, we assumed that all model flies began with goal headings centered within the danger zone but of variable strengths. The structure of flies' exploration is at first dictated by their initial visual map and goal heading. Over time, the goal heading strengthens, which in turn structures the flies' actions and helps to strengthen the visual map (Figure 6D1). Eventually, both the goal heading and the visual map lock onto one of the two safe zones (Figure 6D2). Different flies undergo different experiences depending on the dynamics of their HD representation and on the actions they select based on this representation. However, on average, model flies strengthen their goal headings while shifting them toward safety, which leads to higher PI scores (Figure 6D3).

When examined across individuals, some model flies exhibited learning dynamics that were representative of the ensemble average and quickly converged to the safe zone (Figures 6E and 6F, top row), while others were much slower to learn (Figure 6F, bottom row). We hypothesized that this slower learning might arise from bump jumps that were not coherently aligned with the evolving goal heading, such that jumps would frequently move the bump away from, rather than toward, the goal heading. To test this, we measured the evolving alignment between the most stable compass heading, where active compass neurons experience the least inhibition from active ring neurons, and the goal heading (Figure 6C). We found that a consistently good alignment between the two headings tended

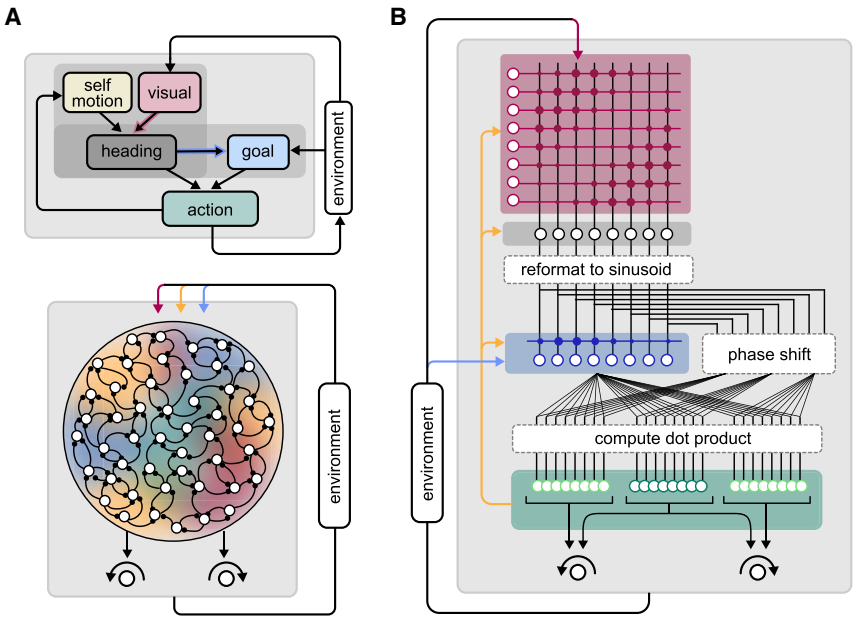


Figure 7. A modular circuit architecture enforces the form of a behavioral policy coupled to evolving internal representations

(A) Recurrent neural network models typically comprise a large ensemble of identical units (circles) connected via plastic synapses. When trained on a task with multiple different inputs (colors), these networks tend to generate highly distributed representations in which individual units exhibit mixed selectivity for several inputs.

(B) In contrast to the diffuse plasticity and distributed representations in (A), our proposed circuit model is highly modular, with distinct cell types carrying and combining internal representations of distinct variables. Inductive biases (dotted boxes) enforce the form of a behavioral policy, which in turn restricts plasticity to a set of synapses onto goal neurons (blue-shaded box). Another set of synapses onto compass neurons is also plastic (red-shaded box).

to allow flies to converge more rapidly onto strong goal headings in the safe zone (Figure 6F, top row). In contrast, prolonged periods of misalignment tended to slow learning, with model flies often getting "stuck" in bad parameter regimes and requiring more time to "escape" to regimes that allowed them to reliably learn safe goal headings (Figure 6F, bottom row). Although we observed this pattern in flies with weak and strong initial goal headings, prolonged periods of misalignment were more likely to occur in model flies that began with weak goal headings. To dissect the relationship between the strength of the goal heading and the visual map onto the compass, we aligned the behavior of all model flies to the time point of their weakest goal heading (Figures 6G and 6H). Following a drop in the strength of the goal heading, model flies tended to suffer from periods of misalignment between the goal heading and most stable compass heading; the weaker the goal heading, the longer and more pronounced was the period of misalignment (Figure 6G2, left). This led to jumps in the HD representation away from the goal heading, which in turn led to higher variance in the updates of the goal heading toward its final value (Figures 6G2 right and S9). These incoherent updates adversely impacted performance by slowing the evolution of PI scores (Figure 6G3).

To better compare these model predictions to the performance of laser-trained flies in our behavioral paradigm, we repeated the simulations shown in Figures 6D–6H, this time varying both the strength and location of the initial goal heading. When we separated real flies into two groups based on the initial strength of their preferred arena heading—a proxy for the initial strength of their goal headings—they showed a similar behavioral evolution to model flies (Figures 6I and S10). Both model and real flies that began with weaker inferred goal headings tended to maintain significantly weaker goal headings across trials (Figure 6I, top row), were slower to shift their goal headings toward the center of the safe zone (Figure 6I, middle row), and were slower to increase their PI scores (Figure 6I, bottom row).

In summary, flies' individual actions determine how the learning process evolves. However, starting with a strong goal heading—even if inaccurate—may structure behavior to enable a faster remapping of the visual environment onto the compass; a faster determination of good headings within the environment; and as a consequence of more stable visual mappings aligned with stronger goal headings, a faster and more accurate updating of internal goals.

DISCUSSION

We sought to understand how animals simultaneously map their surroundings and tether goals to these surroundings during early experience in a novel environment. We modified a well-established operant learning paradigm for *Drosophila*,[16,27] and we used this to explore how flies' behavior is shaped by co-evolving internal representations of compass and goal headings. We first established that flies' learning performance depends on an internal HD representation carried by their compass neurons. We then used calcium imaging in a symmetric visual environment to reveal how this internal HD representation evolves over time in a new setting and how this in turn might impact the formation of an internal goal heading and the flexible modification of behavior toward this goal heading. By combining these physiological and behavioral results with RL algorithms and circuit modeling that were informed by previous CX studies,[24,36–38,44,48,52,65–68,71,74–76] we suggest how evolutionarily conserved circuits in the fly central brain could efficiently guide behavior relative to a changing goal heading and thereby enable rapid behavioral changes in new settings. In contrast to the distributed representations that are typically learned in large, recurrent neural networks (Figure 7A),[77] our proposed architecture is modular and highly structured to optimally implement key computations, with specific cell types carrying, modifying, and combining distinct internal representations to guide

behavior (Figure 7B). In new settings, these internal representations evolve alongside one another to impact behavior, which in turn impacts how animals sample their surroundings and update these internal representations. Our results show how hardwired, modular architectures can speed up this learning process and how the interdependence between these evolving representations could shape individual variability in learning.

Neural underpinnings of rapid learning

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 toward safety. Our results suggest that flies' rapid learning relies on strong inductive biases[25] that enforce the *form* of a behavioral policy and thereby dictate flies' sampling strategy. This policy is implemented in the projection patterns of neurons in the CX,[24,67] which compare current and goal headings to drive motor output.[72,73] This policy assumes the existence of a single goal heading, and it efficiently directs the fly toward such a heading. During learning, the fly can then use experience at one heading to non-locally update actions at all headings, rather than learning associations at each heading individually. In the language of RL, this amounts to scaling and shifting an entire state-action function, rather that learning individual state-action associations. This type of inductive bias reduces the possible scenarios that are explored during learning and can thereby accelerate the learning process when these scenarios are compatible with the learning task.[78,79] Recent RL studies have explored how such inductive biases might be constructed by learning common features across different learning tasks,[80,81] a process known as learning to learn.[82] Here, we show how inductive biases that are likely learned over evolutionary timescales can be inferred directly from an animal's behavior. The ability to rapidly exploit these inductive biases, in this case by modifying a single goal heading, relies on faster-timescale learning. Because these inductive biases guide how flies sample their surroundings, the resulting behavioral structure can help anchor and rapidly update the mapping of the visual world onto an internal compass heading, which can accelerate the learning of the goal heading derived from the compass heading. The evolution of and interaction between these representations alter the degree of exploitative behavior via the weakening and strengthening of the goal heading. Whether and to what extent this faster-timescale learning could modify the form of the behavioral policy, for example, by further increasing the exploitative component through increased training, remain unknown. This could be combined with behavioral state information, for example, about walking versus flight, to switch the specific actions that are controlled through the same behavioral policy.

Learning to control distributions of continuous and discrete movements

We formulated our behavioral policy in terms of two different modes of flight behavior: fixations and saccades. These modes, which are exhibited by both tethered[55] and freely flying[83] flies, are characterized by distinct kinematic properties, and they necessitate both continuous and discrete control[55,84] (additional resources). We incorporated these modes into our behavioral policy, and we used observed variability in kinematic parameters to infer the parametric form and control parameters of this policy. We then used optimal RL algorithms[53] to specify how these control parameters should change based on experience. This approach bears similarities to recent studies proposing that learning operates on low-dimensional generative parameters that control distributions of movements, rather than on the higher-dimensional space of all possible movements.[85–87] This reduction in the dimensionality of parameter space is one factor that accelerates learning; the inductive biases described above further accelerate learning by enforcing additional structure in these parameters across headings. We showed that this structure is pre-built into how untrained flies sample their surroundings, a strategy that might facilitate dispersal in the absence of explicit goals.[56] Indeed, these same patterns of structured behavior resemble those observed in tethered flies responding to innately attractive or aversive visual features[84]; note that the size of saccades was also shown to vary based on angular distance from a "goal" object, something that was less striking in our setting (Figure S6). Although it might seem optimal to steer toward and maintain a single goal heading, rather than probabilistically bias saccades toward this heading, using such a default behavioral strategy would likely be too predictable to avoid predation[88] and would minimize exploration. Indeed, many animals, including flies, show stochasticity in their actions when behaving freely in dynamic settings.[88,89]

Comparisons to previous models

In constructing our proposed circuit architecture, we incorporated only those details of CX anatomy and physiology that we believe to be essential for understanding circuit function in our task. Previous 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.[67,68,71–74] In contrast to some of these models (for example, Goulard et al.[68]), our model does not rely on the HD representation to maintain a consistent relationship to the visual scene; rather, it allows the HD and goal representations to co-evolve, a key aspect of early learning in a novel environment. Additionally, in most of these models, walking insects generate turns of the correct size to reach goal headings, with any deviations attributed to fixed noise. While this may be appropriate in the context of homing behavior,[67,68] our model assumes that the brain can dynamically adjust the degree of exploitative behavior about the goal heading, through an intrinsically probabilistic policy, and that CX neurons control parameters of this distribution. We find that this approach better captures the finer-grained aspects of flies' behavior in our experiments. Recent studies carried out in parallel to ours have highlighted the role of PFL3 neurons[72,73] and PFL2 neurons[73] in walking flies performing menotaxis. Although it is not yet known how these neuron types function during learned behavior in flight, these experimental results are largely consistent with the circuit model we propose. Our proposed mechanisms for how putative goal neurons might operate and for how goal information might be stored, updated, and read out do not yet have strong support in available physiological and anatomical data. However, 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.[71] More recently, experiments involving "FC2A" neurons, whose connectivity and morphology[24] led us to propose that they could carry a goal signal to the PFLs,[90,91] have shown that they may indeed play such a role[72] (see additional resources for further discussion of our model relative to existing physiological data).

A reinterpretation of previous fly visual learning studies

Our results suggest a reinterpretation of many visual learning studies in tethered flying flies.[27,32,35,92] In these studies, the symmetry of the visual environment and associated reinforcement was thought to rule out the possibility of heading-based learning; flies were instead thought to directly associate actions with visual features. Our results suggest that flies associate actions with an internal HD representation and that symmetry-driven jumps in this representation produce the observed bimodality in flies' behavior. Because these jumps copy the same behavioral policy at multiple arena headings that are signaled by the same visual patterns, they bypass any limitations that a single-goal-heading-based policy might otherwise impose on the fly's actions. It is possible that such jumps would be less frequent in free flight, where proprioceptive cues likely play a greater role in controlling compass bump dynamics.[52] Notably, neurons that encode environmental symmetries are not unique to insects: neurons in the retrosplenial cortex of freely behaving rodents also display such responses[93,94] and are also thought to arise via Hebbian plasticity.[95] Such symmetries are not specific to synthetic scenes; natural scenes have repeating visual features that can alter the relationship between the HD representation and the visual scene.[44] More broadly, it is not known whether flies can learn multiple distinct goal headings, nor to what extent they can form more complex policies beyond the 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.[96] In addition to the CX-based learning we study here, it is possible that this and other spatial navigation behaviors may also rely on associations made in the mushroom body,[20,97–103] another brain region that has been linked to operant visual learning in tethered flying flies.[32,104]

The role of other sensorimotor pathways

Our results suggest that learning in this assay is mediated by a flexible pathway through the CX; however, this might work in concert with direct sensorimotor pathways that instruct reflexive actions, for example, to enable the fly to quickly escape punished zones. Responses to innately attractive or aversive objects[43,105–108] could rely on hardwired visuomotor pathways that recruit banks of feature detectors in the optic foci[109,110] and could instruct stereotyped motor responses based on feature detector inputs.[84] These direct pathways, which might underlie the behavior of compass-neuron-silenced flies that favor high bar stimuli (Figures 1I–1K), might also contribute to the behavior of control flies in our paradigm. However, our silencing experiments suggest that flies' responses under laser training, particularly when that training goes against an innate preference, rely on an indirect pathway that recruits a flexible, compass-neuron-dependent behavioral policy. 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 are not yet known (but see additional resources for a discussion of how reflexive actions could be incorporated within our framework).

Outlook

It has 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.[79] Indeed, inductive biases that exploit such structure may hold the key to animal learning.[26] 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 flexible, and perhaps context-dependent, internal representations 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 near future.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Visual arena
  - Flight visual learning
  - Fly preparation for imaging during flight
  - Two-photon calcium imaging
  - Behavioral analysis
  - Bump analysis
  - Modeling
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES
  - Linking the circuit model to known anatomy
  - Reinforcement learning framework
  - General architecture
  - Flexible policy for maintaining a behavioral preference
  - Fixed-form policy tethered to a single goal heading
  - Circuit implementation of a fixed-form policy

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2024.04.036.

ACKNOWLEDGMENTS

We thank the following people: Jason Wittenbach, Daniel Barabasi, Parvez Ahammad, and Alice Wang for early contributions; Bjorn Brembs, Eyal Gruntman, Michael Reiser, Josh Dudman, John Tuthill, T.J. Florence, Sung Soo Kim, Yoshi Aso, and members of the Jayaraman and Hermundstad labs for insightful input; Sandro Romani, Marcella Noorman, Hannah Haberkern, Dan Turner-Evans, Stephen Thornquist, and Gaby Maimon for feedback on previous versions of the manuscript; and Gudrun Ihrke for support through Project Technical Resources. The FPGA Wingbeat Analyzer was developed with Coleman Technologies's Dan Milkie and Andy Chiu. We are grateful to Janelia

Experimental Technology for technical assistance; Jinyang Liu (LED arena); Steve Sawtelle (D2A converter for Wingbeat Analyzer); and Tanya Tabachnik, Igor Negrashov, and Bill Biddle (fly holder for two-photon imaging). We thank Janelia's Fly Facility for stock support and the Media Facility for special fly food. This work was funded by the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

Experiments, C.D. and R.K. (Kir-related experiments); data processing/initial analysis, C.D.; behavioral analyses, A.M.H. (input, C.D., V.J., and B.K.H.); imaging analyses, V.J. and A.M.H. (input, C.D.); theoretical framework, modeling, A.M.H. (conceptual input, V.J., B.K.H., and C.D.); circuit implementation, all authors (with B.K.H. particularly contributing to behavioral policy implementation); interpreting results, all authors; writing, A.M.H. and V.J. (editing, B.K.H. and C.D.).

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 3, 2023
Revised: January 16, 2024
Accepted: April 30, 2024
Published: May 24, 2024

REFERENCES

1. Dupre, C., and Yuste, R. (2017). Non-overlapping neural networks in Hydra vulgaris. Curr. Biol. *27*, 1085–1097. https://doi.org/10.1016/j.cub.2017.02.049.
2. Wadhams, G.H., and Armitage, J.P. (2004). Making sense of it all: bacterial chemotaxis. Nat. Rev. Mol. Cell Biol. *5*, 1024–1037. https://doi.org/10.1038/nrm1524.
3. Huston, S.J., and Jayaraman, V. (2011). Studying sensorimotor integration in insects. Curr. Opin. Neurobiol. *21*, 527–534. https://doi.org/10.1016/j.conb.2011.05.030.
4. Calhoun, A.J., and Murthy, M. (2017). Quantifying behavior to solve sensorimotor transformations: advances from worms and flies. Curr. Opin. Neurobiol. *46*, 90–98. https://doi.org/10.1016/j.conb.2017.08.006.
5. Crochet, S., Lee, S.-H., and Petersen, C.C.H. (2019). Neural circuits for goal-directed sensorimotor transformations. Trends Neurosci. *42*, 66–77. https://doi.org/10.1016/j.tins.2018.08.011.
6. Pouget, A., and Snyder, L.H. (2000). Computational approaches to sensorimotor transformations. Nat. Neurosci. *3*, 1192–1198. https://doi.org/10.1038/81469.
7. Wolpert, D.M., and Flanagan, J.R. (2016). Computations underlying sensorimotor learning. Curr. Opin. Neurobiol. *37*, 7–11. https://doi.org/10.1016/j.conb.2015.12.003.
8. Knierim, J.J., and Zhang, K. (2012). Attractor dynamics of spatially correlated neural activity in the limbic system. Annu. Rev. Neurosci. *35*, 267–285. https://doi.org/10.1146/annurev-neuro-062111-150351.
9. Finkelstein, A., Las, L., and Ulanovsky, N. (2016). 3-D maps and compasses in the brain. Annu. Rev. Neurosci. *39*, 171–196. https://doi.org/10.1146/annurev-neuro-070815-013831.
10. Tolman, E.C., and Honzik, C.H. (1930). Introduction and removal of reward, and maze performance in rats. Univ. California Publications Psychol. *4*, 257–275.
11. Huber, D., Gutnisky, D.A., Peron, S., O'Connor, D.H., Wiegert, J.S., Tian, L., Oertner, T.G., Looger, L.L., and Svoboda, K. (2012). Multiple dynamic representations in the motor cortex during sensorimotor learning. Nature *484*, 473–478. https://doi.org/10.1038/nature11039.
12. Poort, J., Khan, A.G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., Bauza, M., Sahani, M., Keller, G.B., Mrsic-Flogel, T.D., et al. (2015). Learning enhances sensory and multiple non-sensory representations in primary visual cortex. Neuron *86*, 1478–1490. https://doi.org/10.1016/j.neuron.2015.05.037.
13. Peters, A.J., Lee, J., Hedrick, N.G., O'Neil, K., and Komiyama, T. (2017). Reorganization of corticospinal output during motor learning. Nat. Neurosci. *20*, 1133–1141. https://doi.org/10.1038/nn.4596.
14. Coddington, L.T., and Dudman, J.T. (2019). Learning from action: reconsidering movement signaling in midbrain dopamine neuron activity. Neuron *104*, 63–77. https://doi.org/10.1016/j.neuron.2019.08.036.
15. Kuchibhotla, K.V., Hindmarsh Sten, T., Papadoyannis, E.S., Elnozahy, S., Fogelson, K.A., Kumar, R., Boubenec, Y., Holland, P.C., Ostojic, S., and Froemke, R.C. (2019). Dissociating task acquisition from expression during learning reveals latent knowledge. Nat. Commun. *10*, 2151. https://doi.org/10.1038/s41467-019-10089-0.
16. Brembs, B., and Heisenberg, M. (2000). The operant and the classical in conditioned orientation of Drosophila melanogaster at the flight simulator. Learn. Mem. *7*, 104–115. https://doi.org/10.1101/lm.7.2.104.
17. Strauss, R. (2002). The central complex and the genetic dissection of locomotor behaviour. Curr. Opin. Neurobiol. *12*, 633–638. https://doi.org/10.1016/s0959-4388(02)00385-9.
18. Pfeiffer, K., and Homberg, U. (2014). Organization and functional roles of the central complex in the insect brain. Annu. Rev. Entomol. *59*, 165–184. https://doi.org/10.1146/annurev-ento-011613-162031.
19. Turner-Evans, D.B., and Jayaraman, V. (2016). The insect central complex. Curr. Biol. *26*, R453–R457. https://doi.org/10.1016/j.cub.2016.04.006.
20. Webb, B., and Wystrach, A. (2016). Neural mechanisms of insect navigation. Curr. Opin. Insect Sci. *15*, 27–39. https://doi.org/10.1016/j.cois.2016.02.011.
21. Varga, A.G., Kathman, N.D., Martin, J.P., Guo, P., and Ritzmann, R.E. (2017). Spatial navigation and the central complex: sensory acquisition, orientation, and motor control. Front. Behav. Neurosci. *11*, 4. https://doi.org/10.3389/fnbeh.2017.00004.
22. Heinze, S. (2017). Unraveling the neural basis of insect navigation. Curr. Opin. Insect Sci. *24*, 58–67. https://doi.org/10.1016/j.cois.2017.09.001.
23. Honkanen, A., Adden, A., da Silva Freitas, J., and Heinze, S. (2019). The insect central complex and the neural basis of navigational strategies. J. Exp. Biol. *222*, jeb188854. https://doi.org/10.1242/jeb.188854.
24. Hulse, B.K., Haberkern, H., Franconville, R., Turner-Evans, D., Takemura, S.Y., Wolff, T., Noorman, M., Dreher, M., Dan, C., Parekh, R., et al. (2021). A connectome of the *Drosophila* central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. eLife *10*, e66039. https://doi.org/10.7554/eLife.66039.
25. Mitchell, T.M. (1980). The Need for Biases in Learning Generalizations. Rutgers CS tech report CBM-TR-117.
26. Zador, A.M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. Nat. Commun. *10*, 3770. https://doi.org/10.1038/s41467-019-11786-6.
27. Wolf, R., and Heisenberg, M. (1991). Basic organization of operant behavior as revealed in Drosophila flight orientation. J. Comp. Physiol. A *169*, 699–705. https://doi.org/10.1007/BF00194898.
28. Götz, K.G. (1987). Course-control, metabolism and wing interference during ultralong tethered flight in Drosophila melanogaster. J. Exp. Biol. *128*, 35–46. https://doi.org/10.1242/jeb.128.1.35.
29. Reiser, M.B., and Dickinson, M.H. (2008). A modular display system for insect behavioral neuroscience. J. Neurosci. Methods *167*, 127–139. https://doi.org/10.1016/j.jneumeth.2007.07.019.
30. Brembs, B., and Heisenberg, M. (2001). Conditioning with compound stimuli in Drosophila melanogaster in the flight simulator. J. Exp. Biol. *204*, 2849–2859. https://doi.org/10.1242/jeb.204.16.2849.
31. Heisenberg, M., Wolf, R., and Brembs, B. (2001). Flexibility in a single behavioral variable of Drosophila. Learn. Mem. *8*, 1–10. https://doi.org/10.1101/lm.8.1.1.

32. Tang, S., and Guo, A. (2001). Choice behavior of Drosophila facing contradictory visual cues. Science *294*, 1543–1547. https://doi.org/10.1126/science.1058237.

33. Tammero, L.F., and Dickinson, M.H. (2002). The influence of visual landscape on the free flight behavior of the fruit fly Drosophila melanogaster. J. Exp. Biol. *205*, 327–343. https://doi.org/10.1242/jeb.205.3.327.

34. Frye, M.A., and Dickinson, M.H. (2007). Visual edge orientation shapes free-flight behavior in Drosophila. Fly *1*, 153–154. https://doi.org/10.4161/fly.4563.

35. Liu, G., Seiler, H., Wen, A., Zars, T., Ito, K., Wolf, R., Heisenberg, M., and Liu, L. (2006). Distinct memory traces for two visual features in the Drosophila brain. Nature *439*, 551–556. https://doi.org/10.1038/nature04381.

36. Seelig, J.D., and Jayaraman, V. (2015). Neural dynamics for landmark orientation and angular path integration. Nature *521*, 186–191. https://doi.org/10.1038/nature14446.

37. Giraldo, Y.M., Leitch, K.J., Ros, I.G., Warren, T.L., Weir, P.T., and Dickinson, M.H. Sun navigation requires compass neurons in Drosophila. Curr. Biol. 28, 2845–2852.e4. https://doi.org/10.1016/j.cub.2018.07.002.

38. Green, J., Vijayan, V., Mussells Pires, P., Adachi, A., and Maimon, G. (2019). A neural heading estimate is compared with an internal goal to guide oriented navigation. Nat. Neurosci. *22*, 1460–1468. https://doi.org/10.1038/s41593-019-0444-x.

39. Haberkern, H., Chitnis, S.S., Hubbard, P.M., Goulet, T., Hermundstad, A.M., and Jayaraman, V. (2022). Maintaining a stable head direction representation in naturalistic visual environments. Preprint at bioRxiv. https://www.biorxiv.org/content/10.1101/2022.05.17.492284v1.

40. Turner-Evans, D.B., Jensen, K.T., Ali, S., Paterson, T., Sheridan, A., Ray, R.P., Wolff, T., Lauritzen, J.S., Rubin, G.M., Bock, D.D., and Jayaraman, V. (2020). The neuroanatomical ultrastructure and function of a biological ring attractor. Neuron *108*, 145–163.e10.

41. Neuser, K., Triphan, T., Mronz, M., Poeck, B., and Strauss, R. (2008). Analysis of a spatial orientation memory in Drosophila. Nature *453*, 1244–1247. https://doi.org/10.1038/nature07003.

42. Guo, C., Du, Y., Yuan, D., Li, M., Gong, H., Gong, Z., and Liu, L. (2014). A conditioned visual orientation requires the ellipsoid body in Drosophila. Learn. Mem. *22*, 56–63. https://doi.org/10.1101/lm.036863.114.

43. Maimon, G., Straw, A.D., and Dickinson, M.H. (2008). A simple vision-based algorithm for decision making in flying Drosophila. Curr. Biol. *18*, 464–470. https://doi.org/10.1016/j.cub.2008.02.054.

44. Kim, S.S., Hermundstad, A.M., Romani, S., Abbott, L.F., and Jayaraman, V. (2019). Generation of stable heading representations in diverse visual scenes. Nature *576*, 126–131. https://doi.org/10.1038/s41586-019-1767-1.

45. Seelig, J.D., and Jayaraman, V. (2013). Feature detection and orientation tuning in the Drosophila central complex. Nature *503*, 262–266. https://doi.org/10.1038/nature12601.

46. Omoto, J.J., Keleş, M.F., Nguyen, B.M., Bolanos, C., Lovick, J.K., Frye, M.A., and Hartenstein, V. (2017). Visual input to the Drosophila central complex by developmentally and functionally distinct neuronal populations. Curr. Biol. *27*, 1098–1110. https://doi.org/10.1016/j.cub.2017.02.063.

47. Sun, Y., Nern, A., Franconville, R., Dana, H., Schreiter, E.R., Looger, L.L., Svoboda, K., Kim, D.S., Hermundstad, A.M., and Jayaraman, V. (2017). Neural signatures of dynamic stimulus selection in Drosophila. Nat. Neurosci. *20*, 1104–1113. https://doi.org/10.1038/nn.4581.

48. Fisher, Y.E., Lu, J., D'Alessandro, I., and Wilson, R.I. (2019). Sensorimotor experience remaps visual input to a heading-direction network. Nature *576*, 121–125. https://doi.org/10.1038/s41586-019-1772-4.

49. Kottler, B., Faville, R., Bridi, J.C., and Hirth, F. (2019). Inverse control of turning behavior by dopamine D1 receptor signaling in columnar and ring neurons of the central complex in Drosophila. Curr. Biol. *29*, 567–577.e6. https://doi.org/10.1016/j.cub.2019.01.017.

50. Frighetto, G., Zordan, M.A., Castiello, U., Megighian, A., and Martin, J.-R. (2022). Dopamine modulation of Drosophila ellipsoid body neurons, a nod to the mammalian basal ganglia. Front. Physiol. *13*, 849142. https://doi.org/10.3389/fphys.2022.849142.

51. Fisher, Y.E., Marquis, M., D'Alessandro, I., and Wilson, R.I. (2022). Dopamine promotes head direction plasticity during orienting movements. Nature *612*, 316–322. https://doi.org/10.1038/s41586-022-05485-4.

52. Beetz, M.J., Kraus, C., Franzke, M., Dreyer, D., Strube-Bloss, M.F., Rössler, W., Warrant, E.J., Merlin, C., and El Jundi, B. (2022). Flight-induced compass representation in the monarch butterfly heading network. Curr. Biol. *32*, 338–349.e5. https://doi.org/10.1016/j.cub.2021.11.009.

53. Sutton, R.S., and Barto, A.G. (2018). Reinforcement Learning: an Introduction, Second Edition (The MIT Press).

54. Wolf, R., and Heisenberg, M. (1980). On the fine structure of yaw torque in visual flight orientation of Drosophila melanogaster. J. Comp. Physiol. *140*, 69–80. https://doi.org/10.1007/BF00613749.

55. Muijres, F.T., Elzinga, M.J., Iwasaki, N.A., and Dickinson, M.H. (2015). Body saccades of Drosophila consist of stereotyped banked turns. J. Exp. Biol. *218*, 864–875. https://doi.org/10.1242/jeb.114280.

56. Warren, T.L., Weir, P.T.W., and Dickinson, M.H. (2018). Flying Drosophila melanogaster maintain arbitrary but stable headings relative to the angle of polarized light. J. Exp. Biol. *221*, jeb177550. https://doi.org/10.1242/jeb.177550.

57. Haberkern, H., Basnak, M.A., Ahanonu, B., Schauder, D., Cohen, J.D., Bolstad, M., Bruns, C., and Jayaraman, V. (2019). Visually guided behavior and optogenetically induced learning in head-fixed flies exploring a virtual landscape. Curr. Biol. *29*, 1647–1659.e8. https://doi.org/10.1016/j.cub.2019.04.033.

58. Mathejczyk, T.F., and Wernet, M.F. (2019). Heading choices of flying Drosophila under changing angles of polarized light. Sci. Rep. *9*, 16773. https://doi.org/10.1038/s41598-019-53330-y.

59. Williams, C.B. (1957). Insect migration. Annu. Rev. Entomol. *2*, 163–180. https://doi.org/10.1146/annurev.en.02.010157.001115.

60. Coyne, J.A., Boussy, I.A., Prout, T., Bryant, S.H., Jones, J.S., and Moore, J.A. (1982). Long-distance migration of Drosophila. Am. Nat. *119*, 589–595. https://doi.org/10.1086/283936.

61. Wehner, R. (1984). Astronavigation in insects. Annu. Rev. Entomol. *29*, 277–298. https://doi.org/10.1146/annurev.en.29.010184.001425.

62. Weir, P.T., and Dickinson, M.H. (2012). Flying Drosophila orient to sky polarization. Curr. Biol. *22*, 21–27. https://doi.org/10.1016/j.cub.2011.11.026.

63. Dickinson, M.H. (2014). Death valley, Drosophila, and the Devonian toolkit. Annu. Rev. Entomol. *59*, 51–72. https://doi.org/10.1146/annurev-ento-011613-162041.

64. Leitch, K.J., Ponce, F.V., Dickson, W.B., van Breugel, F., and Dickinson, M.H. (2021). The long-distance flight behavior of Drosophila supports an agent-based model for wind-assisted dispersal in insects. Proc. Natl. Acad. Sci. USA *118*, e2013342118. https://doi.org/10.1073/pnas.2013342118.

65. Hartmann, G., and Wehner, R. (1995). The ant's path integration system: a neural architecture. Biol. Cybern. *73*, 483–497.

66. Wittmann, T., and Schwegler, H. (1995). Path integration — a network model. Biol. Cybern. *73*, 569–575. https://doi.org/10.1007/BF00199549.

67. Stone, T., Webb, B., Adden, A., Weddig, N.B., Honkanen, A., Templin, R., Wcislo, W., Scimeca, L., Warrant, E., and Heinze, S. (2017). An anatomically constrained model for path integration in the bee brain. Curr. Biol. *27*, 3069–3085.e11. https://doi.org/10.1016/j.cub.2017.08.052.

68. Goulard, R., Buehlmann, C., Niven, J.E., Graham, P., and Webb, B. (Sept. 2021). A unified mechanism for innate and learned visual landmark guidance in the insect central complex. PLoS Comp. Biol. *17*, e1009383. https://doi.org/10.1371/journal.pcbi.1009383.

69. Touretzky, D.S., Redish, A.D., and Wan, H.S. (Nov. 1993). Neural representation of space Using sinusoidal arrays. Neural Comput. *5*, 869–884. https://doi.org/10.1162/neco.1993.5.6.869.

70. Lyu, C., Abbott, L.F., and Maimon, G. (2022). Building an allocentric travelling direction signal via vector computation. Nature *601*, 92–97. https://doi.org/10.1038/s41586-021-04067-0.

71. Matheson, A.M.M., Lanz, A.J., Medina, A.M., Licata, A.M., Currier, T.A., Syed, M.H., and Nagel, K.I. (2022). A neural circuit for wind-guided olfactory navigation. Nat. Commun. *13*, 4613. https://doi.org/10.1038/s41467-022-32247-7.

72. Pires, P.M., Abbott, L., and Maimon, G. (2022). Converting an allocentric goal into an egocentric steering signal. Preprint at bioRxiv. https://doi.org/10.1101/2022.11.10.516026.

73. Westeinde, E.A., Kellogg, E., Dawson, P.M., Lu, J., Hamburg, L., Midler, B., Druckmann, S., and Wilson, R.I. (2024). Transforming a head direction signal into a goal-oriented steering command. Nature *626*, 819–826. https://doi.org/10.1038/s41586-024-07039-2.

74. Rayshubskiy, A., Holtz, S.L., D'Alessandro, I., Li, A.A., Vanderbeck, Q.X., Haber, I.S., Gibb, P.W., and Wilson, R.I. (2020). Neural circuit mechanisms for steering control in walking Drosophila. Preprint at bioRxiv. https://doi.org/10.1101/2020.04.04.024703v2.

75. Kim, S.S., Rouault, H., Druckmann, S., and Jayaraman, V. (2017). Ring attractor dynamics in the Drosophila central brain. Science *356*, 849–853. https://doi.org/10.1126/science.aal4835.

76. Cope, A.J., Sabo, C., Vasilaki, E., Barron, A.B., and Marshall, J.A.R. (2017). A computational model of the integration of landmarks and motion in the insect central complex. PLoS One *12*, e0172325. https://doi.org/10.1371/journal.pone.0172325.

77. Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. *35*, 1798–1828. https://doi.org/10.1109/TPAMI.2013.50.

78. Bishop, C.M. (2006). Pattern Recognition and Machine Learning, Second Edition (Springer-Verlag).

79. Botvinick, M., Ritter, S., Wang, J.X., Kurth-Nelson, Z., Blundell, C., and Hassabis, D. (2019). Reinforcement learning, fast and slow. Trends Cogn. Sci. *23*, 408–422. https://doi.org/10.1016/j.tics.2019.02.006.

80. Wang, J.X., Kurth-Nelson, Z., Tirumala, D., Soyer, H., Leibo, J.Z., Munos, R., Blundell, C., Kumaran, D., and Botvinick, M. (2016). Learning to Reinforcement Learn. Preprint at arXiv. https://doi.org/10.48550/arXiv.1611.05763.

81. Duan, Y., Schulman, J., Chen, X., Bartlett, P.L., Sutskever, I., and Abbeel, P. (2016). RL2: Fast Reinforcement Learning via Slow Reinforcement Learning. Preprint at arXiv. https://doi.org/10.48550/arXiv.1611.02779.

82. Harlow, H.F. (1949). The formation of learning sets. Psychol. Rev. *56*, 51–65. https://doi.org/10.1037/h0062474.

83. Collett, T.S., and Land, M.F. (1975). Visual control of flight behaviour in the hoverfly Syritta pipiens L. J. Comp. Physiol. *99*, 1–66. https://doi.org/10.1007/BF01464710.

84. Mongeau, J.-M., and Frye, M.A. (2017). Drosophila spatiotemporally integrates visual signals to control saccades. Curr. Biol. *27*, 2901–2914.e2. https://doi.org/10.1016/j.cub.2017.08.035.

85. Yttri, E.A., and Dudman, J.T. (2016). Opponent and bidirectional control of movement velocity in the basal ganglia. Nature *533*, 402–406. https://doi.org/10.1038/nature17639.

86. Zhou, B., Hofmann, D., Pinkoviezky, I., Sober, S.J., and Nemenman, I. (2018). Chance, long tails, and inference in a non-Gaussian, Bayesian theory of vocal learning in songbirds. Proc. Natl. Acad. Sci. USA *115*, E8538–E8546. https://doi.org/10.1073/pnas.1713020115.

87. Jiang, W.-C., Xu, S., and Dudman, J.T. (2022). Construction of a hippocampal cognitive map depends upon spatial context. Nat. Neurosci. *25*, 1693–1705. https://doi.org/10.1038/s41593-022-01201-7.

88. Bolton, A.D., Haesemeyer, M., Jordi, J., Schaechtle, U., Saad, F.A., Mansinghka, V.K., Tenenbaum, J.B., and Engert, F. (2019). Elements of a stochastic 3D prediction engine in larval zebrafish prey capture. eLife *8*, e51975. https://doi.org/10.7554/eLife.51975.

89. Demir, M., Kadakia, N., Anderson, H.D., Clark, D.A., and Emonet, T. (2020). Walking *Drosophila* navigate complex plumes using stochastic decisions biased by the timing of odor encounters. eLife *9*, e57524. https://doi.org/10.7554/eLife.57524.

90. Dan, C., Hulse, B.K., Jayaraman, V., and Hermundstad, A.M. (2021). Flexible control of behavioral variability mediated by an internal representation of head direction. Preprint at bioRxiv. https://doi.org/10.1101/2021.08.18.456004v1.

91. Dan, C., Kappagantula, R., Hulse, B.K., Jayaraman, V., and Hermundstad, A.M. (2022). Flexible control of behavioral variability mediated by an internal representation of head direction. Preprint at bioRxiv. https://doi.org/10.1101/2021.08.18.456004v2.

92. Dill, M., Wolf, R., and Heisenberg, M. (1995). Behavioral analysis of Drosophila: landmark learning in the flight simulator. Learn. Mem. *2*, 152–160. https://doi.org/10.1101/lm.2.3-4.152.

93. Jacob, P.-Y., Casali, G., Spieser, L., Page, H., Overington, D., and Jeffery, K. (2017). An independent, landmark-dominated head-direction signal in dysgranular retrosplenial cortex. Nat. Neurosci. *20*, 173–175. https://doi.org/10.1038/nn.4465.

94. Zhang, N., Grieves, R.M., and Jeffery, K.J. (2022). Environment symmetry drives a multidirectional code in rat retrosplenial cortex. J. Neurosci. *42*, 9227–9241. https://doi.org/10.1523/JNEUROSCI.0619-22.2022.

95. Page, H.J.I., and Jeffery, K.J. (July 2018). Landmark-based updating of the head direction system by retrosplenial cortex: A computational model. Front. Cell. Neurosci. *12*, 191. https://doi.org/10.3389/fncel.2018.00191.

96. Ofstad, T.A., Zuker, C.S., and Reiser, M.B. (2011). Visual place learning in Drosophila melanogaster. Nature *474*, 204–207. https://doi.org/10.1038/nature10131.

97. Mizunami, M., Weibrecht, J.M., and Strausfeld, N.J. (1998). Mushroom bodies of the cockroach: their participation in place memory. J. Comp. Neurol. *402*, 520–537. https://doi.org/10.1002/(SICI)1096-9861(19981228)402:4<520::AID-CNE6>3.0.CO;2-K.

98. Ardin, P., Peng, F., Mangan, M., Lagogiannis, K., and Webb, B. (2016). Using an insect mushroom body circuit to encode route memory in complex natural environments. PLoS Comput. Biol. *12*, e1004683. https://doi.org/10.1371/journal.pcbi.1004683.

99. Collett, M., and Collett, T.S. (2018). How does the insect central complex use mushroom body output for steering? Curr. Biol. *28*, R733–R734. https://doi.org/10.1016/j.cub.2018.05.060.

100. Buehlmann, C., Wozniak, B., Goulard, R., Webb, B., Graham, P., and Niven, J.E. (2020). Mushroom bodies are required for learned visual navigation, but not for innate visual behavior, in ants. Curr. Biol. *30*, 3438–3443.e2. https://doi.org/10.1016/j.cub.2020.07.013.

101. Sun, X., Yue, S., and Mangan, M. (2020). A decentralised neural model explaining optimal integration of navigational strategies in insects. eLife *9*, e54026. https://doi.org/10.7554/eLife.54026.

102. Kamhi, J.F., Barron, A.B., and Narendra, A. (2020). Vertical lobes of the mushroom bodies are essential for view-based navigation in Australian Myrmecia ants. Curr. Biol. *30*, 3432–3437.e3. https://doi.org/10.1016/j.cub.2020.06.030.

103. Bennett, J.E.M., Philippides, A., and Nowotny, T. (2021). Learning with reinforcement prediction errors in a model of the Drosophila mushroom body. Nat. Commun. *12*, 2569. https://doi.org/10.1038/s41467-021-22592-4.

104. Liu, Q., Yang, X., Tian, J., Gao, Z., Wang, M., Li, Y., and Guo, A. (2016). Gap junction networks in mushroom bodies participate in visual learning and memory in Drosophila. eLife *5*, e13238. https://doi.org/10.7554/eLife.13238.

105. Reichardt, W., and Wenking, H. (Aug. 1969). Optical detection and fixation of objects by fixed flying flies. Naturwissenschaften *56*, 424–425. https://doi.org/10.1007/BF00593644.

106. Wehner, R. (1972). Spontaneous pattern preferences of Drosophila melanogaster to black areas in various parts of the visual field. J. Insect Physiol. *18*, 1531–1543. https://doi.org/10.1016/0022-1910(72)90232-6.

107. Horn, E. (1978). The mechanism of object fixation and its relation to spontaneous pattern preferences in Drosophila melanogaster. Biol. Cybern. *31*, 145–158. https://doi.org/10.1007/BF00337000.

108. Grabowska, M.J., Steeves, J., Alpay, J., Van De Poll, M., Ertekin, D., and van Swinderen, B. (2018). Innate visual preferences and behavioral flexibility in Drosophila. J. Exp. Biol. *221*, jeb185918. https://doi.org/10.1242/jeb.185918.

109. Panser, K., Tirian, L., Schulze, F., Villalba, S., Jefferis, G.S.X.E., Bühler, K., and Straw, A.D. (2016). Automatic segmentation of Drosophila neural compartments using GAL4 expression data reveals novel visual pathways. Curr. Biol. *26*, 1943–1954. https://doi.org/10.1016/j.cub.2016.05.052.

110. Klapoetke, N.C., Nern, A., Peek, M.Y., Rogers, E.M., Breads, P., Rubin, G.M., Reiser, M.B., and Card, G.M. (2017). Ultra-selective looming detection from radial motion opponency. Nature *551*, 237–241. https://doi.org/10.1038/nature24626.

111. Scheffer, L.K., Xu, C.S., Januszewski, M., Lu, Z., Takemura, S.Y., Hayworth, K.J., Huang, G.B., Shinomiya, K., Maitlin-Shepard, J., Berg, S., et al. (2020). A connectome and analysis of the adult *Drosophila* central brain. eLife *9*, e57443. https://doi.org/10.7554/eLife.57443.

112. Wolff, T., and Rubin, G.M. (2018). Neuroarchitecture of the Drosophila central complex: A catalog of nodulus and asymmetrical body neurons and a revision of the protocerebral bridge catalog. J. Comp. Neurol. *526*, 2585–2611. https://doi.org/10.1002/cne.24512.

113. Guo, A., Li, L., Xia, S.Z., Feng, C.H., Wolf, R., and Heisenberg, M. (1996). Conditioned visual flight orientation in Drosophila: dependence on age, practice, and diet. Learn. Mem. *3*, 49–59. https://doi.org/10.1101/lm.3.1.49.

114. Wang, J.W., Wong, A.M., Flores, J., Vosshall, L.B., and Axel, R. (2003). Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. Cell *112*, 271–282.

115. Seelig, J.D., Chiappe, M.E., Lott, G.K., Dutta, A., Osborne, J.E., Reiser, M.B., and Jayaraman, V. (2010). Two-photon calcium imaging from head-fixed Drosophila during optomotor walking behavior. Nat. Methods *7*, 535–540. https://doi.org/10.1038/nmeth.1468.

116. Hubert, M., and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. Comp. Stat. Data Anal. *52*, 5186–5201. https://doi.org/10.1016/j.csda.2007.11.008.

117. Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. J. Stat. Software *31*, 1–21.

118. Turner-Evans, D.B., Wegener, S., Rouault, H., Franconville, R., Wolff, T., Seelig, J.D., Druckmann, S., and Jayaraman, V. (2017). Angular velocity integration in a fly heading circuit. eLife *6*, e23496. https://doi.org/10.7554/eLife.23496.

119. Shiozaki, H.M., and Kazama, H. (2017). Parallel encoding of recent visual experience and self-motion during navigation in Drosophila. Nat. Neurosci. *20*, 1395–1403. https://doi.org/10.1038/nn.4628.

120. Hardcastle, B.J., Omoto, J.J., Kandimalla, P., Nguyen, B.M., Keleş, M.F., Boyd, N.K., Hartenstein, V., and Frye, M.A. (2021). A visual pathway for skylight polarization processing in Drosophila. eLife *10*, e63225. https://doi.org/10.7554/eLife.63225.

121. Heinze, S., and Reppert, S.M. (2011). Sun compass integration of skylight cues in migratory monarch butterflies. Neuron *69*, 345–358. https://doi.org/10.1016/j.neuron.2010.12.025.

122. Vitzthum, H., Müller, M., and Homberg, U. (2002). Neurons of the central complex of the locust Schistocerca gregaria are sensitive to polarized light. J. Neurosci. *22*, 1114–1125. https://doi.org/10.1523/JNEUROSCI.22-03-01114.2002.

123. Okubo, T.S., Patella, P., D'Alessandro, I., and Wilson, R.I. (2020). A neural network for wind-guided compass navigation. Neuron *107*, 924–940.e18. https://doi.org/10.1016/j.neuron.2020.06.022.

124. Hanesch, U., Fischbach, K.-F., and Heisenberg, M. (1989). Neuronal architecture of the central complex in Drosophila melanogaster. Cell Tissue Res. *257*, 343–366. https://doi.org/10.1007/BF00261838.

125. Isaacman-Beck, J., Paik, K.C., Wienecke, C.F.R., Yang, H.H., Fisher, Y.E., Wang, I.E., Ishida, I.G., Maimon, G., Wilson, R.I., and Clandinin, T.R. (2020). SPARC enables genetic manipulation of precise proportions of cells. Nat. Neurosci. *23*, 1168–1175. https://doi.org/10.1038/s41593-020-0668-9.

126. Green, J., and Maimon, G. (2018). Building a heading signal from anatomically defined neuron types in the Drosophila central complex. Curr. Opin. Neurobiol. *52*, 156–164. https://doi.org/10.1016/j.conb.2018.06.010.

127. Skaggs, W.E., Knierim, J.J., Kudrimoti, H.S., and McNaughton, B.L. (1995). A model of the neural basis of the rat's sense of direction. Adv. Neural Inf. Process. Syst. *7*, 173–180.

128. Kuntz, S., Poeck, B., and Strauss, R. (2017). Visual working memory requires permissive and instructive NO/cGMP signaling at presynapses in the Drosophila central brain. Curr. Biol. *27*, 613–623. https://doi.org/10.1016/j.cub.2016.12.056.

129. Liang, X., Ho, M.C.W., Zhang, Y., Li, Y., Wu, M.N., Holy, T.E., and Taghert, P.H. (2019). Morning and evening circadian pacemakers independently drive premotor centers via a specific dopamine relay. Neuron *102*, 843–857.e4. https://doi.org/10.1016/j.neuron.2019.03.028.

130. Franconville, R., Beron, C., and Jayaraman, V. (2018). Building a functional connectome of the *Drosophila* central complex. eLife *7*, e37017. https://doi.org/10.7554/eLife.37017.

131. Green, J., Adachi, A., Shah, K.K., Hirokawa, J.D., Magani, P.S., and Maimon, G. (2017). A neural circuit architecture for angular integration in Drosophila. Nature *546*, 101–106. https://doi.org/10.1038/nature22343.

132. Lin, C.Y., Chuang, C.C., Hua, T.E., Chen, C.C., Dickson, B.J., Greenspan, R.J., and Chiang, A.S. (2013). A comprehensive wiring diagram of the protocerebral bridge for visual information processing in the Drosophila brain. Cell Rep. *3*, 1739–1753. https://doi.org/10.1016/j.celrep.2013.04.022.

133. Wolff, T., Iyer, N.A., and Rubin, G.M. (2015). Neuroarchitecture and neuroanatomy of the Drosophila central complex: a GAL4-based dissection of protocerebral bridge neurons and circuits. J. Comp. Neurol. *523*, 997–1037. https://doi.org/10.1002/cne.23705.

134. Currier, T.A., Matheson, A.M., and Nagel, K.I. (2020). Encoding and control of orientation to airflow by a set of Drosophila fan-shaped body neurons. eLife *9*, e61510. https://doi.org/10.7554/eLife.61510.

135. Lu, J., Behbahani, A.H., Hamburg, L., Westeinde, E.A., Dawson, P.M., Lyu, C., Maimon, G., Dickinson, M.H., Druckmann, S., and Wilson, R.I. (2022). Transforming representations of movement from body- to world-centric space. Nature *601*, 98–104. https://doi.org/10.1038/s41586-021-04191-x.

136. Müller, J., Nawrot, M., Menzel, R., and Landgraf, T. (2018). A neural network model for familiarity and context learning during honeybee foraging flights. Biol. Cybern. *112*, 113–126. https://doi.org/10.1007/s00422-017-0732-z.

137. Zhu, L., Mangan, M., and Webb, B. (2020). Spatio-temporal memory for navigation in a mushroom body model. Preprint at bioRxiv. https://doi.org/10.1101/2020.10.27.356535.

138. Hu, W., Peng, Y., Sun, J., Zhang, F., Zhang, X., Wang, L., Li, Q., and Zhong, Y. (2018). Fan-shaped body neurons in the Drosophila brain regulate both innate and conditioned nociceptive avoidance. Cell Rep. *24*, 1573–1584. https://doi.org/10.1016/j.celrep.2018.07.028.

139. Claridge-Chang, A., Roorda, R.D., Vrontou, E., Sjulson, L., Li, H., Hirsh, J., and Miesenböck, G. (2009). Writing memories with light-addressable

reinforcement circuitry. Cell *139*, 405–415. https://doi.org/10.1016/j.cell.2009.08.034.

140. Aso, Y., and Rubin, G.M. (2016). Dopaminergic neurons write and update memories with cell-type-specific rules. eLife *5*, e16135. https://doi.org/10.7554/eLife.16135.
141. Cohn, R., Morantte, I., and Ruta, V. (2015). Coordinated and compartmentalized neuromodulation shapes sensory processing in Drosophila. Cell *163*, 1742–1755. https://doi.org/10.1016/j.cell.2015.11.019.
142. Cognigni, P., Felsenberg, J., and Waddell, S. (2018). Do the right thing: neural network mechanisms of memory formation, expression and update in Drosophila. Curr. Opin. Neurobiol. *49*, 51–58. https://doi.org/10.1016/j.conb.2017.12.002.
143. Siju, K.P., Štih, V., Aimon, S., Gjorgjieva, J., Portugues, R., and Grunwald Kadow, I.C. (2020). Valence and state-dependent population coding in dopaminergic neurons in the fly mushroom body. Curr. Biol. *30*, 2104–2115.e4. https://doi.org/10.1016/j.cub.2020.04.037.
144. Liu, Q., Liu, S., Kodama, L., Driscoll, M.R., and Wu, M.N. (2012). Two dopaminergic neurons signal to the dorsal fan-shaped body to promote wakefulness in Drosophila. Curr. Biol. *22*, 2114–2123. https://doi.org/10.1016/j.cub.2012.09.008.
145. Ueno, T., Tomita, J., Tanimoto, H., Endo, K., Ito, K., Kume, S., and Kume, K. (2012). Identification of a dopamine pathway that regulates sleep and arousal in Drosophila. Nat. Neurosci. *15*, 1516–1523. https://doi.org/10.1038/nn.3238.
146. Cachope, R., Mateo, Y., Mathur, B.N., Irving, J., Wang, H.-L., Morales, M., Lovinger, D.M., and Cheer, J.F. (2012). Selective activation of cholinergic interneurons enhances accumbal phasic dopamine release: setting the tone for reward processing. Cell Rep. *2*, 33–41. https://doi.org/10.1016/j.celrep.2012.05.011.
147. Sulzer, D., Cragg, S.J., and Rice, M.E. (2016). Striatal dopamine neurotransmission: regulation of release and uptake. Basal Ganglia *6*, 123–148. https://doi.org/10.1016/j.baga.2016.02.001.
148. Hsu, C.T., and Bhandawat, V. (2016). Organization of descending neurons in Drosophila melanogaster. Sci. Rep. *6*, 20259. https://doi.org/10.1038/srep20259.
149. Namiki, S., Dickinson, M.H., Wong, A.M., Korff, W., and Card, G.M. (2018). The functional organization of descending sensory-motor pathways in *Drosophila*. eLife *7*, e34272. https://doi.org/10.7554/eLife.34272.
150. Cande, J., Namiki, S., Qiu, J., Korff, W., Card, G.M., Shaevitz, J.W., Stern, D.L., and Berman, G.J. (2018). Optogenetic dissection of descending behavioral control in *Drosophila*. eLife *7*, e34275. https://doi.org/10.7554/eLife.34275.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Two-photon imaging experiments	This paper	https://doi.org/10.25378/janelia.25655817
Behavioral experiments	This paper	https://doi.org/10.25378/janelia.25655817
FIBSEM data	Scheffer et al.[111]	https://neuprint.janelia.org/
Experimental models: Organisms/strains		
Drosophila: Wild Type Berlin (WTB)	Wolf and Heisenberg[27]	N/A
Drosophila: 11D03AD	Janelia Research Campus	N/A
Drosophila: SS00090	Wolff and Rubin[112]	RRID:BDSC_75849
Drosophila: SS00096	Kim et al.[75]	RRID:BDSC_86861
Drosophila: w+,WTB;;UAS-Kir2.1-EGFP	Janelia Research Campus	N/A
Drosophila: 60D05	Bloomington *Drosophila* Stock Center	RRID:BDSC_39247
Drosophila: w+; 20xUAS-GCaMP6f [su(Hw)attP5; attP2,VK00005] (WTB)	Janelia Research Campus	N/A
Software and algorithms		
FPGA camera-based wingbeat analyzer	This paper	https://doi.org/10.5281/zenodo.11060747
Behavioral analysis and plotting	This paper	https://doi.org/10.5281/zenodo.11223915
Calcium activity analysis and plotting	This paper	https://doi.org/10.5281/zenodo.11223915
Modeling and plotting	This paper	https://doi.org/10.5281/zenodo.11223915
MATLAB	MathWorks	RRID:SCR_001622
LabVIEW	National Instruments	RRID:SCR_014325

RESOURCE AVAILABILITY

Lead contact

Requests for further information should be directed to Ann Hermundstad (hermundstada@janelia.hhmi.org).

Materials availability

This study did not generate any new materials.

Data and code availability

- All behavioral data and calcium imaging data are available at https://janelia.figshare.com/articles/dataset/25655817. The DOI is provided in the key resources table.
- Code for the FPGA camera-based wingbeat analyzer is available at https://github.com/ChuntaoDan/FPGA_WingbeatAnalyzer. Code for all behavioral analyses, calcium imaging analyses, modeling, and plotting is available at https://github.com/HermundstadLab/flyVisualLearning. DOIs are provided in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Parental flies were grown sparsely on Wurzburg food in bottles for at least 6 generations.[113] 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 and 3); 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 2). We used the following SS flies: SS00090[112] and SS00096.[75]

METHOD DETAILS

Visual arena

A blue LED circular arena[29] 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.[27,31] In imaging experiments, visual patterns were displayed on a set of cylindrically arranged blue LED panels (464 nm) extending ±90° in azimuth and ±45° in elevation, and tilted by 26° from horizon towards the fly. Two layers of blue filter (ROSCO #59 INDIGO), an electromagnetic shield and a diffuser were used to cover the LED panels as previously described. The rest of the LED display was covered with black aluminum foil. Only half of the 360° pattern was displayed at any given time because the LED display only spanned ±90° in azimuth. A custom camera-based FPGA wingbeat analyzer was used to measure the fly's intended yaw turning instead of the diode-based wingbeat analyzer.

Flight visual learning

All flies first went through a test trial in a closed-loop environment with a single vertical blue stripe (15° w x 120° h) to assess basic flight performance before operant learning experiments in the horizontal bar environment. This trial lasted between 30 s and 120 s; only the first 30 s were used to assess flight performance in Figure S4.

For operant learning experiments, 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 any stops or without any airpuffs 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[114] 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.[115]

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, LUMPlanFI/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.

Behavioral 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 (Figure S5A). 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 Hz and an upper cutoff frequency of 10 Hz (we will denote this filtered signal as $\tilde{A}^{WB}$). 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 $\tilde{A}^{WB}_{\{t\}}$ had a consistent sign.

A turn that produces a sustained nonzero difference in wingbeat amplitude $\tilde{A}^{WB}$ will lead to changes in the arena heading x in the opposite direction. We used this to define a quantity $s_t = -\tilde{A}^{WB}_t|\Delta x_t|$, 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., $t_{end} : s_{t_{end}} < 0.25 s_{t_{start}}$). The remaining timepoints ($t < t_{start}$ and $t > t_{end}$) 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 during which either (i) the wingbeat frequency f^{WB} dropped below a threshold of $f^{WB}_{min} = 0.01$ (for the upright arena) and $f^{WB}_{min} = 1$ (for the two-photon arena) for any amount of time, or (ii) the turn intersected with a period of time that spanned 500 ms before and 500 ms after an airpuff. We then ranked each remaining turn according to a quantity $r(\text{turn}) = |\langle s_{t_{start}:t_{end}} \rangle (x_{t_{end}} - x_{t_{start}})|$ that combines the average change in wingbeat amplitude $\langle s_{t_{start}:t_{end}} \rangle$ with the total change in arena heading $(x_{t_{end}} - x_{t_{start}})$. 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 $r_{thresh} = Q3 + \exp(3M) * 1.5 * \text{IQR}$. Here, $Q3$ is the third quartile (or 75%) of r, r is the inter-quartile range, and M is a skewness estimated using the med-couple of r.[116] 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. For each distribution, we used the MATLAB function *medcouple.m* (created by Francisco Augusto Alcaraz Garcia) to estimate M.

The set of turns for which $r > r_{thresh}$ were classified as saccades. The periods of time between each saccade were events that we then further classified as either fixations or periods of drift. Before this classification, we first removed any periods of time that intersected with an airpuff event. We then identified events for which f^{WB} dropped below f^{WB}_{min} for more than 30 ms; we removed the period of time that spanned 500 ms before the first drop in f^{WB} and 500 ms after the last drop in f^{WB}. Each remaining portion of the event that exceeded 50 ms was then classified as a fixation if the variance in x_t was below 36 pixels, and drift otherwise (if x_t is Gaussian distributed with a standard deviation of σ, this cutoff ensures that 4σ of the distribution falls within 90°). Portions of the event below this 50 ms threshold were classified as "other" and were retained for analyses of PI scores and residencies.

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 deg/ms) is given by $\dot{\theta} = -50(360/96)\tilde{A}^{WB}$, where the factor 50 converts between wingbeat amplitude and pixels/ms, and the factor (360/96) converts from pixels/ms to °/ms. Saccades were characterized by short durations and large average angular velocities, while fixations were characterized by long durations and low average angular velocities (Figure S5B).

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 Δt_{thresh}. We then evaluated fitting performance for 25 evenly spaced values of Δt_{thresh} between 20 ms 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, $\text{IG}(\Delta t; \mu, \lambda)$, was the best-fitting distribution across a majority of scenarios (trials or flies) for thresholds between 100 ms 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 $\Delta t_{\text{thresh}} = 200$ 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 (Figure S5G). This process yields an inverse Gaussian distribution $P(\Delta t) = \text{IG}(\Delta t; a/\nu, a^2/\eta^2)$ 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 $\text{mean} = \mu$ and variance $\text{var} = \mu^3/\lambda$ of the inverse Gaussian distribution satisfy $\log(\text{var}) = 3\log(\text{mean}) - \lambda$. If the variability in the fit parameters can be explained by changes in ν alone (with fixed η^2 and a), the plot of $\log(\text{var})$ versus $3\log(\text{mean})$ will be well-described by a line of slope 1 and fixed offset $-\lambda = -a^2/\eta^2$. Figure S5H 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 $\lambda_F = 0.61$ (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 $a_F = 0.78$ (thus satisfying $\lambda_F = a_F^2/\eta_F^2 = 0.61$). When we restricted our analysis to fixations that were initiated within the danger zone during the first 60 s 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 Figure S5H). 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 $a_F = 0.78$.

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 $P(|\dot{\theta}|)$, where $|\dot{\theta}|$ 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 $|\Delta\theta| = |\dot{\theta}|\Delta t$. We then characterized the distribution of durations conditioned on speed, $P(\Delta t||\dot{\theta}|)$.

Prior to fitting, we removed saccades whose speeds were below a threshold $|\dot{\theta}|_{\text{thresh}} = 0.1°/\text{ms}$. We used the MATLAB function *allfitdist.m* (created by Mike Sheppard) to fit 16 different parametric distributions to the distribution of speeds $P(|\dot{\theta}|)$ 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 $\text{logn}(|\dot{\theta}|; \varphi, \sigma^2)$, 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 $(N_{\text{CW}} - N_{\text{CCW}})/(N_{\text{CW}} + N_{\text{CCW}})$, 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 (Figure S5D upper), suggesting that flies can adaptively control directional bias of their saccades.

The distribution $P(|\dot{\theta}|)$ specifies the probability of initiating saccades of different speeds. For saccades of a given speed $|\dot{\theta}|$, there is significant variability in their duration (Figure S5F). To characterize this variability, we considered 36 equally spaced values of $|\dot{\theta}|$ between 0.25°/ms and 2°/ms. For each value of $|\dot{\theta}|$, we used the MATLAB function *allfitdist.m* to determine the parametric function that best fit the distribution of saccade durations $P(\Delta t||\dot{\theta}|)$ 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 $a_S = 0.56$ but variable drift rate ν_S (Figure S5I), analogous to fixations. In this case, the drift rate increased nonlinearly with the speed $|\dot{\theta}|$ (inset of Figure S5I); we used a least-squares fit to determine the parameters of the best-fitting sigmoid $f(|\dot{\theta}|) = f_M/[1 + \exp(-k(|\dot{\theta}| - |\dot{\theta}|_0))] - f_0$; these where given by $f_M = 2.32$, $k = 7.55$, $|\dot{\theta}|_0 = 0.35$, and $f_0 = -1.26$. Thus, for a saccade initiated with speed $|\dot{\theta}|$, the duration can be generated via a drift diffusion process with a velocity-dependent drift rate $\nu_S = f(|\dot{\theta}|)$, and fixed values of $\eta_S = 1$ and $a_S = 0.56$.

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 (Figure S5J). 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 circuit model used 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 $\sigma_S = 0.54$ (this approximates the distribution of saccade sizes fit 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 = 300 ms (this approximates the median saccade duration measured across all flies, trials, and datasets). In the below table summarizes these choices.

Parameter values used in RL and circuit models			
General policy parameters			
T_{tot}	240	total simulation time in s	duration of two trials
t_S	320	duration of saccade in ms	median duration in data
δt	0.001	timescale of drift diffusion process in s	sampling rate of behavioral data
η_F	1	spread of drift diffusion process	estimated from distribution
a_F	0.79	threshold for drift diffusion process	of fixation durations
φ_S	3.89	parameters of lognormal distribution	estimated from distribution
σ_S	0.54	over saccade sizes (in deg)	of saccade sizes
Parameters for flexible policy			
k_F	1	sensitivity of drift rate	chosen for illustration
$f_{0,F}$	−0.01	sets minimum and maximum scale	constrains avg fixation duration
$f_{M,F}$	10	of drift rate	between 100 ms and 120 s
k_S	1	sensitivity of saccade probability	chosen for illustration
$f_{0,S}$	−0.01	sets minimum and maximum scale	constrains saccade probability
$f_{M,S}$	0.98	of saccade probability	between 0.01 and 0.99
n	16	number of von Mises functions	matched to EB tiling
κ	8	concentration of von Mises functions	–
Parameters for setpoint (“fixed-form”) policy			
G_S	0.9	gain of saccades (controls heading-dependence)	chosen for illustration
B_S	0	baseline direction of saccades	–
G_F	0.8	gain of fixations (controls heading-dependence)	chosen for illustration
B_F	0.05	baseline duration of fixations	–
G_J	0.7	gain of bump jump (controls heading-dependence)	chosen for illustration
B_J	0.05	baseline probability of jump	–
$\Delta\theta_J$	180	size of bump jump in deg	matched to data
Parameters for circuit implementation of setpoint policy			
N	32	discretization of heading space	chosen for illustration
κ	π	concentration of von Mises functions	chosen for illustration
α_C	0.05	learning rate of compass weights	chosen for illustration
α_G	0.001	learning rate of goal weights	–

Data selection

For all analyses shown in the main text, we used only those trials for which the fly exhibited at least 30 s of continuous flight (these periods of continuous flight were defined using the set of saccades, fixations, drift, and ‘‘other’’ periods that we extracted from our segmentation; see method details section Partitioning behavior into fixations and saccades). For flies and trials that met this selection criterion, we kept all additional periods of flight that exceeded 3.788 ms (defined by the 75th percentile of fixation durations computed across all flies and all trials within the no-laser control dataset). 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 Δt_{thresh} = 200 ms, and those saccades that exceeded an average angular velocity of $|\dot{\theta}|_{\text{thresh}} = 0.1^\circ/\text{ms}$.

Assessing bearings relative to a single stripe

Figure S4 shows the strength and orientation of flies’ bearings relative to a single stripe. These were computed by taking the circular mean of the fictive heading directions using the first 30 s of the single-stripe trial (see method details section flight visual learning) using the MATLAB Circular Statistics Toolbox.[117]

Measuring the strength of behavioral preference

Figure 1 shows the strength of behavioral preferences, measured with respect to the arena preferences of individual flies (Figures 1F, 6I, and S2) or with respect to the safe zone (Figures 1G, 1H, 1J, and 1K). In both cases, we computed the preference strength using the performance index (PI) score. When measuring the strength of preference for safety, PI scores were computed as $\text{PI}_{\text{safe}} = (T_{\text{safe}} - T_{\text{danger}})/(T_{\text{safe}} + T_{\text{danger}})$, where T_{safe} and T_{danger} denote the total time spent in safe and danger zones,

respectively.[27] When measuring the strength of individual preferences, PI scores were computed analogously: $PI_{pref} = (T_{pref} - T_{anti-pref})/(T_{pref} + T_{anti-pref})$, where the "preferred" and "anti-preferred" zones were defined analogously to "safe" and "danger" zones (i.e., arena headings were partitioned into two sets of preferred and anti-preferred zones, each spanning 90°), but were centered around the preferred arena heading measured in a given trial (see method details section Aligning to individual preference for details about determining the preferred arena heading). The left panels of Figures 1F–1H, 1J, 1K, and S2 show the strength of preference averaged across flies on each trial; the right panels show the strength of preference for data averaged across trials 1-2 and trials 8-9. To perform this average, we used the fraction of time that the fly spent flying on each trial to perform a weighted average of the PI scores across trials 1 and 2, and separately across trials 8 and 9; we excluded flies that did not meet our data selection criteria for all four trials. The upper right panel of Figure 6I shows the strength of preference for groups of flies that began with preferences that were stronger or weaker than the median initial preference strength.

Aligning to individual preference

Figures 1F, 3F, and 6I report features of the behavioral data computed after aligning the data to the arena preference of individual flies. 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 Figures 3F and S6, we first selected the set of saccades and fixations taken by each fly within naive trials (trials 1–2), and within training/probe trials (trials 3-9). 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 11 pixels, centered on a given pixel. For each fly, we 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 Figure 3F, 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. Figures 3F and S6 display the mean and standard error in each of these quantities, computed across flies; in Figure S6, data is collapsed across symmetric regions of the arena and is aligned to either the center of the safe zone or the preferred arena heading of individual flies. For measurements of WBF and saccade size/velocity shown in Figure S6, we scaled each heading-dependent quantity by its fly-specific average value (measured across headings for the given set of trials) before averaging across flies. For measurements of saccade size/velocity/WBF, we separately performed this scaling for CW versus CCW saccades. We then computed the average between CW and CCW saccade properties before averaging across flies.

Measuring distance to safety

The middle righthand panel of Figure 6I 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 method details section aligning to individual preference. We then computed the minimum angular distance between this preferred heading and the center of the closest safe zone. Figure 6I shows the average and standard error of this distance for laser-trained flies.

Bump analysis

Computing calcium transients

For volume imaging of EPG GCaMP activity (Figure 2H), 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 F_0. The resulting time series were Savitzky-Golay filtered with a 3rd order polynomial over 5 frames.

Rather than compute the population vector average (PVA), as in past work,[36,40,44,75,118] we focused here on tracking peaks in EPG population activity ('bump position' or 'compass 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 Figure 2H. 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

The lefthand panel of Figure 2I shows the offsets between the bump and the visual scene for individual flies. To cluster these offsets, we used the built-in MATLAB function *kde.m* (with 256 mesh points) to perform kernel density estimation of the distribution of offsets for each fly on each trial (note that for previous versions of the analyses, we used the version of *kde.m* created by Zdravko Botev, available at https://www.mathworks.com/matlabcentral/fileexchange/14034-kernel-density-estimator). We then used the built-in 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 then 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. The lefthand panel of Figure 2I shows the fraction of time spent at different offsets for individual flies on individual trials.

Characterizing the number and angular separation of offsets

The righthand panels of Figure 2I show 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 upper right panel of Figure 2I. 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 Figure 2I.

Computing HD tuning curves

Figure 2J 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 $\Delta F / F_0$ of each wedge for each given scene orientation (see method details section Computing calcium transients). The HD tuning curves to the right of the main panel of Figure 2J 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

The righthand panel of Figure 5D shows the probability of bump jumps as a function of their location within the EB, measured relative to an inferred goal heading. 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. We took advantage of our previous results (shown in the lower right panel of Figure 2I) 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 compass heading within the EB (the "goal" location). 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 compass heading in the EB. The righthand panel of Figure 5D 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. We computed this conditional jump probability for each fly individually, using data accumulated across trials; we included only those trials for which the bump maintained two different offsets relative to the visual scene, and we included only those flies for which the bump jumped 10 or more times across all trials. We then computed the median conditional jump probability across flies; this is shown as the histogram in Figure 5D. To summarize 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; this is shown as the solid line in Figure 5D.

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 dynamics have been reported in both flying and walking flies in symmetric visual settings in larger arenas as well.[36,44,48]

Modeling

Determining the optimal policy for maintaining a goal heading

Figure 3D 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 additional resources section Reinforcement learning framework (see Algorithm 3); the parameters used in the model are summarized in the above table.

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 additional resources section 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 Equation 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 compass 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 Figure 3D.

Figure 3E illustrates the expected behavioral readout if we couple the optimal policy to an internal representation of heading tethered to a symmetric visual scene. 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.

Summary of compass circuit model

We constructed a circuit model that could account for the tethering of the compass heading to the visual world, which is dictated by a set of plastic weights $\vec{\omega}_C$ from inhibitory ring neurons onto compass neurons that are updated via anti-Hebbian plasticity. We assumed that these weights are updated during saccades based on the velocity of the saccade, the fly's current compass heading during the saccade, and the current view of the visual scene.[44] We used these weights to determine the probability that the HD bump would jump between locations that correspond to symmetric views of the visual scene (see Equation 27 in additional resources section Reinforcement learning framework). Bump jumps were assumed to occur immediately following a saccade. We assumed that these weights could be modified continuously, regardless of the presence or absence of reward/punishment.

Summary of policy circuit model

In addition to the compass circuit model, we constructed a policy circuit model that could implement the form of a behavioral policy, and flexibly modify its parameters. This model is described in detail in additional resources section Reinforcement learning framework. Briefly, we modeled a fly that can fixate and saccade. The duration of its fixations and the directionality of its saccades were 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 were modulated by the strength of the goal heading, and by the fly's current compass heading relative to this goal heading. The location and strength of the goal heading was determined by a set of plastic goal weights $\vec{\omega}_G$ that could change over time based on the fly's current compass heading and current level of reinforcement. In contrast to the compass weights (see method details section summary of compass circuit model), we assumed that the goal weights, $\vec{\omega}_G$, could only be modified in the presence of reward/punishment. Algorithms 3–4 show how we implemented this model.

Summary of circuit model simulations

In Figures 2, 5, and 6, we coupled the compass and policy circuit models (described above), such that the heading output of the compass circuit was used to modify goal and select actions within the policy circuit. We assumed that the compass heading was maintained as a von Mises activity profile in the compass circuit, and that it was transformed into a sinusoidal activity profile before passing into the policy circuit; we did not explicitly model this circuit transformation. In Figures 2 and 5, we assumed that the goal heading was fixed and that there was no reward/punishment; we used these simulations to study the evolution of the compass weights. In Figure 6, we mimicked the experimental setup and 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 goal weights were co-evolving over time. Each training period consisted of a series of iterations, each consisting of a single saccade and a single fixation. Compass weights were iteratively updated at each angular increment (each value of θ_i) during each saccade; compass weights were not updated during fixations. When rewards/punishment was being delivered, we subsampled periods of fixation into 100 ms increments, and iteratively updated the goal weights for each increment; goal weights were not updated during saccades.

The duration of fixations and directionality of saccades were determined by the current goal weights (see Equation 35 in additional resources section Reinforcement learning framework for more details), and the sizes of saccades were sampled from a lognormal distribution with parameters $\varphi_S = 3.89$ and $\sigma_S = 0.54$ (approximating the values that were fit from data). We assumed that all saccades lasted a fixed duration of 300 ms (approximating 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.

Flexible mapping of visual scenes onto the compass heading

Figures 2 and 5 Illustrate the 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 (Figures 2B and 2C), 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 (Figures 2D, 2F, 5C, and 5D), 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 Equation 39 for the update rule). We updated these weights during saccades and in proportion to the squared velocity of each saccade, assuming that velocity was constant through the duration of each saccade.

All weight matrices in Figures 2B–2D were generated for a single simulation that evolved from a randomly initialized weight matrix; in these simulations, we did not allow the bump to jump, even for symmetric scenes. In Figure 2F, we froze the final weight matrix from 2d, and we used it to generate a short simulation of the heading and arena trajectories. We used the policy circuit model to generate fixations and saccades and assuming a fixed, sinusoidal goal profile normalized between 0 and 1 (see method details section summary of policy circuit model for details). In this simulation, we allowed 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.

In Figure 5C, we randomly initialized a weight matrix and allowed it to stabilize in an asymmetric visual scene (again using the policy circuit model to generate fixations and saccades, and again assuming a fixed, sinusoidal goal profile normalized between 0 and 1). We then changed the visual scene to a symmetric one, and simulated the evolution of the weight matrix in the new scene. The heatmaps in Figure 5C shows one such simulation; the panels below the heatmaps illustrate the probability that the bump would jump from any given orientation in the EB (as described above, this was determined by comparing the net inhibition from active ring neurons at the orientation of the compass heading versus the symmetric orientation).

In Figures 5D and 5E, we repeated the simulation shown in Figure 5C while varying the strength of the goal heading. In all cases, the goal profile was sinusoidal, normalized between 0 and a max amplitude A. We fixed the orientation of the goal profile but varied A, using values of [0.2, 0.4, 0.6, 0.8, 1]. For each amplitude, we simulated 200 model flies; the lefthand panels of Figures 5D and 5E report averages across groups of model flies with the same goal strength. We additionally analyzed the relationship between the final strength of the visual map and the final strength of behavioral preference (measured analogously to how we measure it in real flies; see method details section Measuring the strength of behavioral preference for details). The upper righthand panel of Figure 5E shows the best linear fit between these quantities for the strongest goal heading, and the R^2 values of this fit for each goal heading. The lower righthand panel of Figure 5E shows the temporal evolution of the behavioral preference for the strongest goal heading, and for two different initial conditions of the visual map.

A fixed-form behavioral policy tethered to a flexible goal heading

Figures 4B–4E 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 (Figures 4C and 4D). When computing this for phase-shifted compass headings (Figure 4E), the phase shift was applied before the multiplication with the goal activity profile. We assumed that the compass weights were fixed and uniquely specified the arena heading.

Figure 4H 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 Equation 39. Figure 4H 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 Equations 33, 34, and 35).

Co-evolution of two learning systems

Figures 6D–6H shows simulations in which both the compass weights and the goal weights are evolving over time. For all simulations, we first initialized the compass weights in an asymmetric environment. 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 chosen to be one of 5 evenly spaced values ([0.2, 0.4, 0.6, 0.8, 1]), and with an orientation that was centered at $\theta = 0$ (i.e., aligned with what would become the center of the danger zone). 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. Figures 6D–6H summarizes the evolution of the compass weights, goal weights, and PI scores over 1000 model flies (200 model flies for each goal strength). To assess the compass and goal weights, we track the orientation and strength of the most stable compass heading (computed as the circular mean of $(1 - P_{\text{jump}}(\theta))$, and described in the text as the orientation and strength of the visual map) and the goal heading (computed as the circular mean of the goal activity profile). 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 vector maps in Figure 6D summarize the effects of learning across all simulations, as viewed through different quantities related to the orientation and strength of the goal heading and visual map. We binned all quantities into 15 evenly spaces bins; goal strengths were binned between 0.1 and 0.8; visual map strengths were binned between 0.1 and 0.6, the distance to safety was binned between 0° and 90°, and the map and goal orientation were binned between 0° and 360°. For each bin in a given vector map, we identified all instances where a simulation fell in the given bin, regardless of when this occurred during the simulation; the heatmaps in each vector map show the number of instances that fell into any given bin. For a given set of instances within a given bin, we determined the change in those instances over the next timestep of the simulation. In other words, for all instances of an initial goal and map strength, we computed the change in strength for each instance over the next timestep, and averaged these changes

across instances. The vector maps show the magnitude and direction of this change, and the are colored by the average PI score of all instances in the bin. Vector maps were generated using the built-in MATLAB function *quiver.m* with scales of 3, 2, and 2 for panels (d-1), (d-2), and (d-3), respectively.

Figure 6E shows the same vector map as in Figure 6D3, but with individual trajectories superimposed. The selected model flies (which are also shown in the top panels of Figure 6F) were selected by randomly drawing one of the top 10 model flies that had the highest average PI scores over the entire course of the simulation. The model flies in the lower panels of Figure 6F were analogously selected by randomly drawing one of the bottom 10 model flies that had the lowest average PI scores over the entire course of the simulation. For each model fly, we reported the fraction of the simulation time needed to reach a goal strength larger than 0.6 and a distance to safety lower than 22.5°.

Figures 6G and 6H tracks the strength of the goal heading, the angular difference (or "misalignment") 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 Figure 6G), 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 Figure 6H. The righthand panels of Figures 6G2 and 6H2 show the average coherence of goal updates over time. For each model fly, we took the final goal heading at the end of the simulation, and we used this to compute the angular difference between the current and final goal heading as a function of time during the simulation. If the goal is shifting coherently toward its final location, we would expect this angular difference to decrease steadily over time. To measure this, we computed the variance in the distribution of angular differences at successive timepoints (i.e., we computed $\mathrm{var}\left(\Delta\theta_{G,t} - \Delta\theta_{G,t-1}\right)$, where $\Delta\theta_G$ is the angular difference between the current and final goal headings). We used $\log_{10}$(variance) as a measure of the degree of incoherence in goal updates.

The simulations in Figures 6G and 6H were conducted identically to those in 6d-h, with one exception: instead of fixing the initial goal location, we randomly sampled it for each model by centering the goal heading on one of 32 randomly selected bins between 0° and 360°. We then partitioned model flies into two groups depending on whether their initial goal strength was larger or smaller than the median initial goal strength. For each group, we measured the average strength of behavioral preference, average distance between the goal heading and the center of the safe zone, and average PI score as a function of time throughout the simulation. Colors used in Figures 6D–6I, S9, and S10 were generated from a set of perceptually uniform colormaps created by Ander Biguri, available at https://www.mathworks.com/matlabcentral/fileexchange/51986-perceptually-uniform-colormaps.

QUANTIFICATION AND STATISTICAL ANALYSIS

For all statistical tests used in the paper, we use stars to denote the following p values: *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; ****$p \leq 0.0001$. All p values greater than 0.05 were denoted not significant (n.s.). In the lefthand panel of Figures 1F–1H, 1J, and 1K, we measured the trial-specific preferences of individual flies, and we reported the significant differences between groups (groups defined below); significance was determined via a two-sided Wilcoxon rank sum test against the null hypothesis that the strengths of individual preferences came from distributions with different medians. Error bars denote ± the standard error of the mean. In the righthand panel of Figures 1F–1H, 1J, and 1K, we compared the strength of individual preferences within early versus late trials. We reported the significant differences within each group compared between early versus late trials; we also reported the significant differences between groups, and the significant differences between groups compared between early trials and between late trials. Significance within groups was measured via a paired, two-sided Wilcoxon signed rank test against the null hypothesis that the difference in strengths of individual preferences between naive and probe trials has a median of 0. Significance between groups was measured via a two-sided Wilcoxon rank sum test against the null hypothesis that the strengths of individual preferences come from distributions with different medians. Significance was determined using the following groups of flies, where n denotes the number of flies in each group: Figures 1F and 1G compare $n = 44$ laser-trained flies versus $n = 40$ no-laser control flies; Figure 1H compares $n = 44$ kir-silenced flies versus $n = 40$ parental control flies; Figures 1J and 1K compare $n = 22$ kir-silenced flies versus $n = 20$ parental control flies.

Figure 3G measures the trial-specific distance to safety of individual flies, and we reported the significant differences between $n = 44$ laser-trained flies and $n = 40$ no-laser control flies. Significance was determined via a two-sided Wilcoxon rank sum test against the null hypothesis that the distance to safety came from continuous distributions with the same medians. Error bars denote ± the standard error of the mean.

The righthand panels of Figure 6I measure the trial-specific strength of individual preference, distance to safety, and preference score of individual flies. We reported the significant differences between $n = 21$ laser-trained flies that began with strong initial preferences versus $n = 21$ laser-trained flies that began with weak initial preferences. Significance was determined via a two-sided Wilcoxon rank sum test against the null hypothesis that scores measured for flies with strong versus weak initial preferences come from continuous distributions with the same medians. Error bars denote ± the standard error of the mean.

ADDITIONAL RESOURCES

Linking the circuit model to known anatomy

Our model (Figure S7A) builds on physiological, behavioral, and connectomic findings from many laboratories, as well as on the many conceptual ideas and models that have been proposed for the CX in recent years. Although our model focuses on the CX, we note that many visual learning behaviors, including those associated with navigation, may also involve the mushroom body.[32,68,100,102] In addition, the task we employ and most other orienting tasks almost certainly recruit more direct sensorimotor pathways as well, for example, when the fly responds to aversive heat in our task. We chose to exclude such parallel pathways in our modeling framework to more thoroughly investigate the dynamics of early learning driven by internal representations. In the CX, several of our modeling assumptions are based on physiological studies of different CX neuron types during visual stimulation and behavior. Importantly, although not all the key features of our circuit model have physiological support, they are all inspired by the known anatomy and connectivity of the CX. However, rather than incorporating all the known details of CX connectivity, we greatly simplified and abstracted the circuit in order to focus on a few key computations that we believe underlie much of the fly's behavior in the visual learning paradigm. We now discuss the many simplifications that we made, and summarize what is known about several neuron types and network motifs that are likely to play a role in relevant CX circuit computations.

The flexible mapping from visual scene to HD representation: Recurrent connections between ring neurons and EPG neurons

The HD system is tethered to its sensory surroundings by multiple ring neuron classes —many with tens of ring neurons each— that carry information about sensory cues, such as visual features,[45–47,119] polarized light patterns,[120–122] and wind direction.[123] Most of these neurons are thought to be GABA-ergic and inhibitory.[124,125] Some visual-feature-sensitive ring neurons have spatiotemporal receptive fields,[47] and are thus likely to be sensitive to how a visual feature moves across the fly's eyes, a factor that we ignored in our model. Sensory ring neurons are connected all-to-all to other ring neurons of the same class, and, in some cases, across classes as well, and most sensory ring neurons receive feedback from the compass (EPG) neurons.[24] These motifs may ensure that the fly's HD representation tethers to the strongest cues available,[24] which we assume to be less relevant in a visual setting with four identical horizontal bars. Plasticity in the synapses between ring and compass neurons has been hypothesized to create a flexible mapping between sensory cues and the HD representation,[76,126] an idea similar to one proposed for the rodent HD system[127] (Figure S7B). Recent experimental results strongly support this idea of plasticity between visual inputs and compass neurons.[44,48] As part of a model proposed in one of these studies,[44] we assumed that this plasticity depends on an inhibitory Hebbian-like rule that relies on correlated activity between visual and compass neurons during saccades, and results in changes in the depth of inhibition that compass neurons receive at different angular orientations in their surroundings.[48] Plasticity in the EB may involve nitric oxide signaling[128] and motor-state-dependent neuromodulation. There are multiple sources of neuromodulation in the EB, including the ExR2 dopaminergic neurons (Figure S7B), which receive inputs in the LAL and project to the BU and EB (see Figure 14 and associated figure supplements in Hulse et al.[24]) and have been linked to circadian changes in locomotion.[129] Recent experimental results show that the ExR2 neurons indeed play a key role in turn-dependent dopaminergic modulation of plasticity during scene mapping, at least in walking flies.[51]

Several experimental studies[36,44,48] have reported variability in the EPG bump's offset relative to the fly's surroundings in symmetric visual settings. Two visually indistinguishable headings are likely to evoke similar ring neuron population activity, making both EB locations corresponding to those headings viable for the bump to occupy. Which one of those locations the bump resides in would depend on the relative strength of the ring-neuron-to-EPG mapping in those two EB locations. In our behavioral paradigm, we found that some EB locations were more likely to feature bump jumps (see right panel of Figure 5D), and that these bump jumps tended to be 180° in magnitude (see lower right panel of Figure 2I), reflecting the symmetry of the scene. Note that the vertical span of some ring neurons' receptive fields may be large enough[45] to evoke responses to horizontal bars at both high and low elevations, making the scene weakly symmetric at 90° from the perspective of those inputs to the compass neurons, and triggering a few 90° bump jumps as well (small peak in bottom right panel Figure 2I). We explicitly modeled the inhibitory interactions from ring neurons onto EPG neurons, and assumed that the relative, experience-dependent strength of the HD representation in 180°-opposite EB positions determines the probability of an EPG bump jump between them. Note that these differences in the strength of summed ring neuron inhibition onto EPG neurons at 180°-opposite EB locations would not necessarily lead to EPG bump amplitude differences at those locations, because of other sources of broad feedback inhibition onto EPG neurons in both the EB and the PB.[24,40,70,130]

Maintaining and updating the HD representation: A ring attractor circuit involving EB and PB neurons

There are ~48 EPG neurons, which each occupy one of 16–18 compartments in the EB and PB.[24,111] In our model, we assumed that the HD representation is carried by 32 EPG-like neurons, each with distinct HD tuning equally spaced across 360°. The dynamics of the HD representation match those produced by ring attractor networks.[40,75,118] These dynamics depend not just on sensory inputs, but also on self-motion input from the PEN_a (Figure S7C) and other columnar neurons that link the EB and PB.[40,118,131] These self-motion inputs are also important for the mapping of visual scenes onto the EPG population.[44] In our model, the HD representation was entirely driven by visual input through the ring neurons, and we ignored recurrent connections involving the PEN and PEG neurons, as well as the intra-EB connections between different EPG neurons. These connections could, in effect, tether the EPG bump more strongly to locations near its current location, reducing the probability of bump jumps even if the relative strength of connections from ring neurons to different EPG neurons make them more likely. Further, recent extracellular recordings from candidate EPG-like

neurons in monarch butterflies suggest that bump jumps may be less likely when animals are allowed to physically turn[52] rather than being in closed-loop VR as in our imaging experiments.

Neurons that reformat the HD bump and link the PB to the FB

Our model used a von Mises function to represent the shape of the EPG population's HD bump in the EB, but assumed that the bump shape becomes sinusoidal as it goes through the PB. Note that we did not explicitly include connections from EPG neurons onto the broadly arborizing PB neurons that are thought to be key to ensure the sinusoidal shape, the Δ7 neurons[24,70] (Figure S7C). Both the EPG neurons and Δ7 neurons contact a wide range of columnar neurons in the PB. Some of these neurons project back to the EB, but most are FB columnar neurons,[24] such as the PFNs and PFRs[124,132,133] (Figure S7D). Many of these FB columnar neurons receive additional inputs in other CX structures.[24,70,71,134,135] In combination with neuron-type-specific anatomical phase shifts in their projection patterns from individual glomeruli in the PB to columns of the FB[24,70,135] (see Figure S7D for illustrative examples), these inputs likely allow the PB-FB columnar neuron types to participate in vector computations that transform the HD representation in different ways.[24,70,135] This may be highly relevant to the computations that flies use when navigating over long distances in natural conditions.[60,64] In such natural settings, flies would likely need to select actions to maintain a particular traveling direction ('goal heading') rather than to maintain a specific head direction ('goal HD'). Recent conceptual insights from the connectome[24] and, in parallel, confirmatory evidence from physiological experiments[70,135] suggest that some FB columnar neurons use translational self-motion cues to transform HD into an explicit representation of traveling direction (heading). At the end of this Supplemental section, we discuss how a potential circuit mechanism to learn and express a heading (traveling direction) preference that is robust to perturbations from wind might be implemented in FB circuitry. However, in a head-fixed preparation in which the fly only controls and receives visual feedback for its angular movements, the distinction between HD and heading is less relevant. Thus, although we use the term 'heading' in the main text and below, our model did not incorporate mechanisms that would be necessary for this flexible behavior to operate in the space of heading rather than HD.

Learning and storing a goal heading: Candidate neurons and circuitry in the FB

Recent studies have proposed simple conceptual models for how goal headings might be stored in the strength of synaptic connections between neurons of the FB.[24,74] An entirely different model for visually guided homing is that heading-dependent views or visual snapshots are stored in the strength of synapses between visually responsive Kenyon cells and mushroom body output neurons (MBONs) in the MB.[98,99,101,136] There is, as yet, only indirect experimental evidence for the MB model.[100,137] However, there is evidence that flexibility in goal headings depends on visual input to ring neurons[40] and on output from EPG neurons.[37,38] Further, this flexibility does not involve changes in the mapping between the visual scene and the EPG HD representation,[38] suggesting that goal headings are stored downstream of the EPG neurons. The behavioral genetics evidence implicating FB neurons in a visual learning task that inspired ours[35] suggest that goal headings may be stored in plastic synapses between neurons in the FB.

We assumed that the goal heading is stored in the strength of synapses from hypothesized tangential motor state neurons to putative columnar 'goal neurons' (Figures S7E and S7F). Such a mechanism would be metabolically efficient, allowing goal neurons to be activated into a goal-heading pattern specifically when the fly is moving, but not otherwise. We further assumed that these synaptic strengths are modified through the action of neuromodulatory tangential FB neurons, which, we assume, deliver reinforcement signals shaped like the current heading bump (Figure S7E). There is already some physiological evidence for reinforcement signals in the FB,[138] although the neuronal players involved are as yet unknown. Prime candidates for such a role are the ventral FB dopaminergic neuron (DAN) types. MB DANs are known to carry reinforcement (and movement) signals and be involved in associative learning in that brain region,[139–143] and a subset of FB DANs —potentially the more ventral FB DANs, such as FB2A, FB4L, FB4M, and FB5H, rather than the dorsal FB DANs[144,145]— may well perform similar functions. Interestingly, most tangential neurons that innervate the FB receive local input from columnar FB neurons —including subtypes of PFN, PFR, *v*Δ and *h*Δ neurons— near their presynaptic sites in the FB.[24] Any excitatory input that an FB DAN receives from a columnar neuron is likely to depolarize much of the tangential neuron's layered arbors, given the short electrotonic distances involved, but we speculate that input from columnar neurons may control local Ca, a mechanism hypothesized to enable presynaptic modulation of DA vesicle release in the mammalian striatum.[146,147] Such a mechanism would ensure that any reinforcement signals that the FB DANs carry would be locally influenced by heading input, as required by our model. In our model, this heading-shaped neuromodulation would act on synapses between tangential motor-state neurons and goal neurons (Figures S7E and S7F). FB DANs as well as other tangential FB neurons send their outputs to other FB tangential neurons and to many columnar neurons, including the *h*Δ, *v*Δ, and FC neurons.[24] Several of these FB tangential neurons receive inputs in the LAL. For example, FB5A neurons receive synaptic inputs in the LAL, including from PFL2 and PFL3 neurons, making them ideal candidates to carry motor state signals (Figure S7F). Although different FB columnar neurons could act as goal neurons in different contexts, recent experiments suggest that FC2A neurons, which we hypothesized as candidates to carry a goal signal,[90,91] may be capable of influencing walking flies' actions in a manner consistent with such a function,[72] while in other experiments, a class of *h*Δ neurons have been suggested to store goal headings.[71]

Implementing a policy of a fixed form: Phase shifts of CX output neurons

A remarkable feature of several CX columnar neurons is the precision of their projection patterns within different CX structures. In particular, most CX columnar neurons that project from the PB to either the EB or the FB show precise phase shifts between their

localized arbors in the different structures. These phase shifts are computed relative to the projected position of the EPG bump in different structures. For example, PEN neurons whose arbors overlap with EPG neurons in the PB project to a location in the EB that is shifted by 45° relative to their input EPGs.[133] In the case of the PEN_a neurons, this phase shift has been proposed to allow self-motion-derived angular velocity input to shift the position of EPG population activity in the EB.[118,131] Our model relies on the projection patterns of the PFL2 and PFL3 neurons, which show phase shifts of 180° and 90° respectively between their PB and FB arbors (Figures S7D and S7F; note also that PFL2 neurons project to both left and right LALs). Our proposal for how these phase shifts might enable action selection is based on ideas proposed in Hulse et al.[24] and, in the case of the PFL3 neurons, also shares similarities with a model that was originally proposed for path integration in the sweat bee.[67] The same idea was used for models in previous versions of this manuscript,[90,91] and related ideas have been used in work from other laboratories.[68,71–73] As shown in Figure 4E, the phase shifts automatically enable a multiplication of a phase-shifted version of the fly's current heading with its goal heading. For the PFL2 neurons, the 180° shift means that the product of this multiplication peaks when the fly is heading in exactly the opposite direction to the goal heading. If activity in the PFL2 neurons modulates drift rate within neurons that control fixation, as we propose, then the result of peak activity would be high drift rate that results in shorter fixation and transitions to saccades. Thus, PFL phase shifts provide a potential mechanism to directly induce actions that would steer the bump towards a goal heading, allowing learning to work in the lower-dimensional space of merely updating the goal weight vector, rather than the space of actions necessary to direct the fly to its goal. In addition to their direct projections onto descending neurons (DNs)[24,74] that send motor commands to the thorax[148–150] (Figure S7F), the PFL2 and PFL3 neurons also converge onto LAL neurons that themselves project onto DNs.[24] These projection patterns justify our modeling assumptions regarding how heading-tethered PFL (action neuron) activity is converted into directional motor commands. Recent experimental data in walking flies, collected in two independent studies carried out in parallel to ours,[72,73] is consistent with many of our assumptions. Both PFL2 and PFL3 activity appear to vary based on the difference between the compass and goal heading. Note that in contrast to our assumption that PFL activity results from the multiplication of goal and current heading activity, Pires et al.[72] find that the spike rate of PFL3 neurons is a nonlinear function of their membrane potential, which tracks the sum of goal and heading activity. Neither of these studies explicitly analyzed how PFL2/PFL3 activity correlates with the probability of the fly turning or the duration of its fixations, but stimulation of left or right PFL3 activity probabilistically triggers turns in walking flies,[72] consistent with our assumptions for flight. PFL2 activity is highest when the fly's current heading is 180° away from its goal, as we assume, and is correlated with turning speed,[73] an effect we see as well, albeit weakly (Figure S4). These results are also consistent with PFL2 neurons controlling the frequency of saccades via the fixational drift rate (with lower drift rates leading to longer fixations and less frequent saccades).

A potential circuit mechanism to learn and travel in the direction of a true goal heading

As we discussed above, our model ignores the distinction between HD and heading for the purposes of modeling the fly's behavior in our paradigm. However, we believe that this paradigm exploits a mechanism that the fly uses during dispersal and long-range navigation in more natural settings.[60,64] The FB's circuits could provide the requisite mechanism for the fly to travel in a specific direction, that is, to learn and then progress towards a goal heading. In essence, to produce appropriate movements, PFL neurons would need to receive FB input that has already accounted for the effects of different head-body angles and perturbations such as wind input through appropriate vector computations.[24,70]

For example, imagine a fly attempting to travel northeast on a windy day. At every moment in time, the fly's total translational velocity ('TV') vector will be the sum of two components: the influence of the wind and the fly's self-generated movement, which we assume is in the same direction as its head direction (i.e., the fly's head direction matches its body direction). That is, $\mathrm{TV}(t) = \mathrm{Wind}(t) + \mathrm{Fly}(t)$. In this case, if the wind were blowing the fly east, the fly would have to fly north with the same velocity as $\mathrm{Wind}(t)$ to maintain a northeast heading. Here we assume that $\mathrm{Fly}(t)$ is the HD bump scaled by the fly's forward flight velocity, and that $\mathrm{TV}(t)$ is the fly's total translational velocity vector, carried by *hΔB* neurons.[70,135] With these two signals, the fly could compute the allocentic wind direction (i.e., $\mathrm{Wind}(t) = \mathrm{TV}(t) - \mathrm{Fly}(t)$) (but see Matheson et al.[71] and Currier et al.[134]). Next, to compute the desired, or 'goal' (G), head direction, the fly would have to subtract the instantaneous wind direction from the goal translation vector (i.e., $\mathrm{HD}_G(t) = \mathrm{TV}_G - \mathrm{Wind}(t)$). Here we assume that TV_G is a vector that is stored in the FB that encodes the fly's desired travel direction. Once the desired head direction ($\mathrm{HD}_G(t)$) has been computed, the PFL2/3 neurons could generate goal-directed motor commands as described above. The complexity of this computation arises because the PFL2/3 neurons are thought to inherit the fly's HD in the PB, which would prevent them from directly comparing the fly's instantaneous TV ($\mathrm{TV}(t)$) to its goal TV (TV_G). Instead, to accurately account for the influence of the wind, the PFL2/3 neurons would have to receive the desired HD vector ($\mathrm{HD}_G(t)$) as input in the FB. In this way, the PFL2/3 neurons could compare the fly's current HD to its goal HD. This conceptual model demonstrates that the fly CX could, in principle, account for external perturbations when selecting appropriate actions. Alternatively, instead of storing a desired travel direction, the fly could store a desired HD. Doing so could allow the fly to set an approximately accurate course under some circumstances. For example, if the wind consistently blew the fly east, maintaining a north head direction would yield an overall northeast travel, on average, but this mechanisms would not allow the fly to account for external perturbations like the wind. Future physiological recordings during behavior are required to assess which of these two classes of goal vectors the fly may be using and in which contexts.

Reinforcement learning framework

In this study, we used a reinforcement learning framework to explore how the fly's behavioral modes, namely fixations and saccades, should be structured as a function of the fly's heading and updated based on the fly's experience. We first segmented behavior into these two modes, and used the variability in these modes to inform the structure and adaptable control parameters of a behavioral policy, as described in method details section Inferring the structure of a behavioral policy. Here, we build an agent that uses this policy to structure its behavior as a function of its orientation within its visual surroundings. We then used reinforcement learning to train the control parameters of this policy.

In the main text, we considered the objective of maintaining a goal heading. We used this objective to explore how behavior might have been structured over evolutionary timescales, and then compared the results of this learning to naive fly behavior. As shown in Figure 3D, this objective gives rise to control parameters that are sinusoidally structured as a function of the fly's compass heading. We showed that these control parameters, when mediated by an internal representation of heading and controlled with respect to an internal goal heading, produce behavior that qualitatively resembles fly behavior. With such a policy that is structured with respect to this internal goal heading, the fly need only shift the location of the goal heading to adapt to new surroundings. We thus refer to this policy as a "fixed-form" policy that specifies a set of actions that are tethered to the fly's internal representation of heading relative to an internal goal heading. Below, we show how reinforcement learning can be used on shorter timescales to shift the location of the goal heading based on experience. Finally, we illustrate how a simplified circuit model can be used to implement this fixed-form policy, and how plasticity in two sets of weights within the circuit can be used to update the goal heading and most stable compass heading over time.

In what follows, we first outline the general form of the policy and the training algorithm. We then show how the control parameters of this policy can be learned using a policy gradient algorithm with function approximation; we use this approach to study how the control parameters should be structured as a function of heading to maintain a preference for a specific visual pattern or a specific heading. We then assume that the structure in these control parameters is built-in to the policy and maintained with respect to a single goal heading, and we show how a policy gradient algorithm can be used to update the location of this goal heading while otherwise maintaining the structure of the policy. Finally, we outline the circuit implementation of this fixed-form policy, and we show how Hebbian-like plasticity can be used to implement the learning process.

General architecture

Policy

We consider a general scenario in which the behavior of an agent (model fly) is governed by a stochastic policy $\pi(\dot{\theta}, \Delta t|\theta)$. This policy determines the probability of maintaining an average angular velocity $\dot{\theta}$ over a duration of time Δt given an initial heading θ, and thereby determines the probability of generating a change in heading $\Delta\theta = \dot{\theta}\Delta t$. We will use Δt to denote the duration of a single action that is sampled from the policy; depending on the scenario, this can correspond to the entire duration of a saccade, the entire duration of a fixation, or the duration of a sampling event within a fixation. The policy is parameterized by a set of fixed parameters $\vec{\beta}$ that do not change over time, and a set of flexible parameters $\vec{\omega}$ that can be modified through experience. For notational simplicity, we will introduce these parameters as they become necessary. When writing a conditional distribution $P(A|B)$, we will explicitly denote the dependence on $\vec{\omega}$ when it exists (i.e., $P(A|B;\vec{\omega})$), and will assume implicit dependence on $\vec{\beta}$ when it exists.

We decompose behavior into two different behavioral modes: fixations ('F'), and saccades ('S'), such that the policy can be written as:

$$\begin{aligned}\pi\left(\dot{\theta}, \Delta t\middle|\theta\right) &= P\left(\dot{\theta}, \Delta t\middle|\theta, F\right)P(F|\theta) + P\left(\dot{\theta}, \Delta t\middle|\theta, S\right)P(S|\theta) \\ &= P\left(\Delta t\middle|\dot{\theta}, \theta, F\right)P\left(\dot{\theta}\middle|\theta, F\right)P(F|\theta) + P\left(\Delta t\middle|\dot{\theta}, \theta, S\right)P\left(\dot{\theta}\middle|\theta, S\right)(1 - P(F|\theta))\end{aligned} \quad \text{(Equation 1)}$$

Here, we have used the constraint that $P(F|\theta) + P(S|\theta) = 1$, and we have incorporated the observation that the duration of each mode can be conditioned on the angular velocity (see method details sections Characterizing fixation properties and Characterizing saccade properties).

Fixation policy

We assume that fixations are generated in a timepoint-by-timepoint manner via a drift diffusion process with integrated signal ξ; when this signal crosses a fixed threshold, the fixation is terminated and a saccade is initiated. The probability of maintaining a fixation thus depends on ξ:

$$\pi\left(\dot{\theta}, \Delta t\middle|\theta, \xi\right) = P\left(\Delta t\middle|\dot{\theta}, \theta, F\right)P\left(\dot{\theta}\middle|\theta, F\right)P(F|\theta, \xi) + P\left(\Delta t\middle|\dot{\theta}, \theta, S\right)P\left(\dot{\theta}\middle|\theta, S\right)(1 - P(F|\theta, \xi)) \quad \text{(Equation 2)}$$

We assume that the integrated signal ξ is updated in time increments of δt, and we assume that there is no change in heading during this time increment. This allows us to define:

$$P\left(\Delta t\middle|\dot{\theta},\theta,F\right) = \delta(\Delta t - \delta t) \quad \text{(Equation 3)}$$

$$P\left(\dot{\theta}\middle|\theta,F\right) = \delta\left(\dot{\theta}\right) \quad \text{(Equation 4)}$$

During fixations, the integrated signal ξ is updated by an amount $\Delta\xi$ that is determined by the drift diffusion process. We model this process with a fixed spread η_F^2 and a heading-dependent drift rate $\nu_F(\theta;\vec{\omega})$ that is parameterized by the flexible parameters $\vec{\omega}$:

$$P(\Delta\xi|\theta,\xi;\vec{\omega}) = \mathcal{N}\left(\nu_F(\theta;\vec{\omega})\delta t,\eta_F^2\delta t\right) \quad \text{(Equation 5)}$$

This update will terminate the fixation and result in a saccade if the net signal $\xi+\Delta\xi$ crosses a fixed threshold a_F. Note that this produces an inverse Gaussian distribution of fixation durations ΔT, with average durations $a_F/\nu_F(\theta;\vec{\omega})$ (where here, we use ΔT to denote the duration of the entire fixational event, computed in increments of δt):

$$P\left(\Delta T\middle|\dot{\theta},\theta,F;\vec{\omega}\right) = \mathrm{IG}\left(\Delta T;a_F\,/\,\nu_F(\theta;\vec{\omega}),a_F^2\,/\,\eta_F^2\right) \quad \text{(Equation 6)}$$

This allows us to write the probability of fixating as:

$$\begin{aligned} P(F|\theta,\xi;\vec{\omega}) &= \int P(F|\Delta\xi,\theta,\xi)P(\Delta\xi|\theta,\xi)d\Delta\xi \\ &= \int^{a_F} P(\Delta\xi|\theta,\xi)d\Delta\xi \\ &= \frac{1}{2}\left(1+\mathrm{erf}\left(\frac{a_F - (\xi+\nu_F(\theta;\vec{\omega})\delta t)}{\sqrt{2\eta_F^2\delta t}}\right)\right) \end{aligned} \quad \text{(Equation 7)}$$

In the presence of heat, one can include second, reflexive drift process that can short-circuit the termination of a fixation (see method details section Characterizing fixation properties for motivation). To illustrate how this could be carried out, we consider a process that is governed by an integrated signal ξ_R that obeys the same dynamics as above, with the same spread η_F^2 but with a fixed drift rate ν_R. A fixation is terminated whenever either ξ or ξ_R cross the fixed threshold a_F. The probability of fixating is thus determined by:

$$P(F|\theta,\xi,\xi_R;\vec{\omega}) = \min\{P(F|\theta,\xi;\vec{\omega}),P(F|\theta,\xi_R)\} \quad \text{(Equation 8)}$$

where

$$P(F|\theta,\xi_R) = \frac{1}{2}\left(1+\mathrm{erf}\left(\frac{a_F - (\xi_R+\nu_R\delta t)\Theta(\mathrm{heat})}{\sqrt{2\eta_F^2\delta t}}\right)\right) \quad \text{(Equation 9)}$$

Here, $\Theta(\mathrm{heat})$ is a heaviside function that takes a value of 1 if there is a perceived heat, and 0 otherwise. Note that in the absence of perceived heat, $P(F|\theta,\xi;\vec{\omega})$ will always be less than $P(F|\theta,\xi_R)$, and will thus determine the probability of fixating through Equation 8.

Saccade policy

We assume that saccades are initiated in a ballistic manner following the termination of a fixation, such that taking a saccade results in an abrupt change in heading $\Delta\theta = \dot{\theta}\Delta t$ over a duration of time Δt.

We assume that the directionality of saccades is controlled via a heading-dependent directional bias $d_S(\theta;\vec{\omega})$ that is parameterized by the flexible parameters $\vec{\omega}$; $d_S(\theta;\vec{\omega})$ specifies the probability of initiating a rightward, or CW, saccade at heading θ. We then assume that the angular speed of a saccade $|\dot{\theta}|$ is drawn from a lognormal distribution with parameters φ_S,σ_S^2. Together, this results in the following distribution over angular velocities $\dot{\theta}$:

$$P\left(\dot{\theta}\middle|\theta,S;\vec{\omega}\right) = \left[d_S(\theta;\vec{\omega})\mathrm{sgn}\left(\dot{\theta}\right)+\left(\frac{1-\mathrm{sgn}\left(\dot{\theta}\right)}{2}\right)\right]\mathrm{logn}\left(\left|\dot{\theta}\right|;\varphi_S,\sigma_S^2\right) \quad \text{(Equation 10)}$$

We further assume that the duration of saccades, analogously to the duration of fixations, can be generated via a drift diffusion process. Here, we assume that the drift rate is not flexible, but depends on the angular speed of the saccade:

$$P\left(\Delta t\middle|\dot{\theta},\theta,S\right) = \mathrm{IG}\left(\Delta t;a_S\,/\,\nu_S\left(\left|\dot{\theta}\right|\right),a_S^2\,/\,\eta_S^2\right) \quad \text{(Equation 11)}$$

where $\nu_S(|\dot{\theta}|)$ is well-captured by a sigmoidal function of $|\dot{\theta}|$ (see method details section Characterizing saccade properties). We can now specify the full policy and its parameter dependence:

$$\pi\left(\dot{\theta}, \Delta t \middle| \theta, \xi; \vec{\omega}\right) = P\left(\Delta t \middle| \dot{\theta}, \theta, F\right) P\left(\dot{\theta} \middle| \theta, F\right) P(F|\theta, \xi; \vec{\omega}) + P\left(\Delta t \middle| \dot{\theta}, \theta, S\right) P\left(\dot{\theta} \middle| \theta, S; \vec{\omega}\right) (1 - P(F|\theta, \xi; \vec{\omega}))$$

$$\begin{aligned} P\left(\Delta t \middle| \dot{\theta}, \theta, F\right) &= \delta(\Delta t - \delta t) \\ P\left(\dot{\theta} \middle| \theta, F\right) &= \delta\left(\dot{\theta}\right) \\ P(F|\theta, \xi; \vec{\omega}) &= \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{a_F - (\xi + \nu_F(\theta; \vec{\omega})\delta t)}{\sqrt{2\eta_F^2 \delta t}}\right)\right) \\ P\left(\Delta t \middle| \dot{\theta}, \theta, S\right) &= \mathrm{IG}\left(\Delta t; a_S \middle/ \nu_S\left(|\dot{\theta}|\right), a_S^2 \middle/ \eta_S^2\right) \\ P\left(\dot{\theta} \middle| \theta, S; \vec{\omega}\right) &= \left[d_S(\theta; \vec{\omega})\mathrm{sgn}\left(\dot{\theta}\right) + \left(\frac{1 - \mathrm{sgn}\left(\dot{\theta}\right)}{2}\right)\right] \mathrm{logn}\left(|\dot{\theta}|; \varphi_S, \sigma_S^2\right) \end{aligned} \quad \text{(Equation 12)}$$

where $\vec{\omega}$ controls the heading dependence in both the duration of fixations (through the drift rate ν_F) and the directional bias of saccades (through d_S). As noted above, the policy depends implicitly on a set of fixed parameters $\vec{\beta} = [\delta t, a_F, \eta_F, \varphi_S, \sigma_S, a_S, \eta_S]$ that controls inflexible aspects of behavior.

Training

We use an online policy-gradient method to iteratively update the flexible policy parameters $\vec{\omega}$ based on the agent's actions and on the outcome of these actions. In this way, the agent (i) samples an action $[\dot{\theta}, \Delta t]$ from its policy based on its current heading θ, (ii) observes the outcome of this action (a change in heading $\Delta\theta = \dot{\theta}\Delta t$, and a heading-dependent sensory response $R(\theta + \dot{\theta}\Delta t)$), and (iii) updates the policy weights $\vec{\omega}$ to modify the probability of taking the same action from the same heading in the future (depending on whether that action led to a good or bad outcome). We use a policy gradient algorithm[53] to update these weights:

$$\begin{aligned} \Delta\vec{\omega} &= R\left(\theta + \dot{\theta}\Delta t\right) \nabla_{\vec{\omega}} \log \pi\left(\dot{\theta}, \Delta t \middle| \theta; \vec{\omega}\right)\Big|_{\theta^*, \dot{\theta}^*, \Delta t^*} \\ &= R\left(\theta + \dot{\theta}\Delta t\right) \frac{\nabla_{\vec{\omega}} \pi\left(\dot{\theta}, \Delta t \middle| \theta; \vec{\omega}\right)}{\pi\left(\dot{\theta}, \Delta t \middle| \theta; \vec{\omega}\right)}\Bigg|_{\theta^*, \dot{\theta}^*, \Delta t^*} \end{aligned} \quad \text{(Equation 13)}$$

where θ^*, $\dot{\theta}^*$, and Δt^* denote specific values of the heading, angular velocity, and duration of an action, respectively. We assume that $R(\theta + \dot{\theta}\Delta t)$ is computed directly from changes in sensory experience, rather than from a comparison between sensory experience and expected value. Note that the sensory response effectively acts as the step size, or learning rate, in the update equation for $\vec{\omega}$. As a result, a strong sensory response can result in quick but coarse updates, whereas as weak sensory response will result in slower but finer updates. The policy gradients can then be computed as follows:

$$\begin{aligned} \nabla_{\vec{\omega}}\pi &= P\left(\Delta t \middle| \dot{\theta}, \theta, F\right) P\left(\dot{\theta} \middle| \theta, F\right) \nabla_{\vec{\omega}} P(F|\theta, \xi; \vec{\omega}) \\ &\quad + P\left(\Delta t \middle| \dot{\theta}, \theta, S\right) \left[(1 - P(F|\theta, \xi; \vec{\omega}))\nabla_{\vec{\omega}} P\left(\dot{\theta} \middle| \theta, S; \vec{\omega}\right) - P\left(\dot{\theta} \middle| \theta, S; \vec{\omega}\right) \nabla_{\vec{\omega}} P(F|\theta, \xi; \vec{\omega})\right] \\ &= \left[P\left(\Delta t \middle| \dot{\theta}, \theta, F\right) P\left(\dot{\theta} \middle| \theta, F\right) - P\left(\Delta t \middle| \dot{\theta}, \theta, S\right) P\left(\dot{\theta} \middle| \theta, S; \vec{\omega}\right)\right] \nabla_{\vec{\omega}} P(F|\theta, \xi; \vec{\omega}) \\ &\quad + \left[P\left(\Delta t \middle| \dot{\theta}, \theta, S\right)(1 - P(F|\theta, \xi; \vec{\omega}))\right] \nabla_{\vec{\omega}} P\left(\dot{\theta} \middle| \theta, S; \vec{\omega}\right) \end{aligned} \quad \text{(Equation 14)}$$

We can use Equation 12 to further simplify the gradients:

$$\nabla_{\vec{\omega}} P(F|\theta, \xi; \vec{\omega}) = -\delta t \mathcal{N}\left(a_F; \xi + \nu_F(\theta; \vec{\omega})\delta t, \eta_F^2 \delta t\right) \nabla_{\vec{\omega}} \nu_F(\theta; \vec{\omega}) \quad \text{(Equation 15)}$$

$$\nabla_{\vec{\omega}} P\left(\dot{\theta} \middle| \theta, S; \vec{\omega}\right) = \mathrm{sgn}\left(\dot{\theta}\right) \mathrm{logn}\left(|\dot{\theta}|; \varphi_S, \sigma_S^2\right) \nabla_{\vec{\omega}} d_S(\theta; \vec{\omega}) \quad \text{(Equation 16)}$$

Further evaluating the policy gradients depends on the form of $\nabla_{\vec{\omega}} \nu_F(\theta; \vec{\omega})$ and $\nabla_{\vec{\omega}} d_S(\theta; \vec{\omega})$, which are determined by the specific implementations that we consider below.

Implementation

Here and in the main text, we consider different variants of this basic framework. The first variant, shown in Figure 3D, considers a "fully flexible" policy in which $\nu_F(\theta;\vec{\omega})$ and $d_S(\theta;\vec{\omega})$ can be modified in a heading-dependent manner via function approximation with a set of weights $\vec{\omega} = [\vec{\omega}_F, \vec{\omega}_S]$. Learning then acts to change the functional form of $\nu_F(\theta;\vec{\omega}_F)$ and $d_S(\theta;\vec{\omega}_S)$ by modifying $\vec{\omega}$. An extreme version of this policy, in which learning acts independently at each of a finely discretized set of headings, is schematized in Figure 3C1. The second variant, schematized in Figure 3C2 and detailed in this SI, considers a "fixed-form" policy in which the heading-dependence in both $\nu_F(\theta;\vec{\omega})$ and $d_S(\theta;\vec{\omega})$ is structured with respect to a "goal heading" θ_G. In this case, the flexible parameters $\vec{\omega} = \theta_G$ specify the goal heading, and learning acts by shifting the goal heading while preserving the structured heading dependence in $\nu_F(\theta;\vec{\omega})$ and $d_S(\theta;\vec{\omega})$. The final variant, shown in Figures 4, 5, and 6, considers a circuit-based implementation of this fixed-form policy that is informed by physiology and connectomic data (see additional resources section links to known anatomy).

In what follows, we outline the assumptions and training algorithms for these model variants. In each variant, we remove the temporal variability in the duration of saccades by assuming that $P(\Delta t|\dot{\theta},\theta,S) = \delta(\Delta t - t_S)$, where t_S is a constant (we will assume t_S = 300 ms, based on the analysis described in method details section Characterizing saccade properties; see table).

Flexible policy for maintaining a behavioral preference

Policy

We approximate the limited angular resolution of the heading representation via a set of radial basis functions $g_i(\theta)$ ($i = 1\ldots n$) that mimic the anatomical tiling of compass neurons in the Ellipsoid Body. Unless otherwise specified, we used $n = 16$ von Mises basis functions that uniformly tiled the range [0°,360°], with concentration factor $\kappa = 8$. With this representation, heading dependence in the drift rate of fixations ν_F and the directional bias of saccades d_S can be achieved by constructing different weighted combinations of these basis functions, with weights $\vec{\omega}_F$ controlling the heading dependence in ν_F, and weights $\vec{\omega}_S$ controlling heading dependence in d_S:

$$\begin{aligned} \nu_F\left(\theta;\vec{\omega}_F\right) &= f\left(\vec{\omega}_F^T\vec{g}(\theta);k_F,f_{0F},f_{MF}\right) \\ d_S\left(\theta;\vec{\omega}_S\right) &= f\left(\vec{\omega}_S^T\vec{g}(\theta);k_S,f_{0S},f_{MS}\right) \end{aligned} \qquad \text{(Equation 17)}$$

where $\vec{\omega}_a^T\vec{g}(\theta)$ ($a \in \{F,S\}$) is a weighted sum over basis functions evaluated at θ. The sigmoidal function $f(x;k,f_0,f_M) = f_M/(1+\exp(-kx)) - f_0$ enforces bounds on the drift rate of fixations and the probability of a rightward saccade, given a set of parameters $[k,f_0,f_M]$ that control the slope, minimum, and maximum values of the sigmoid. We chose the values of these parameters to bound the drift rate between 0.01 and 1.01, and to bound the probability of rightward saccades between 0 and 1.

As noted above, this policy depends implicitly on a set of fixed parameters $\vec{\beta}$, which now includes the additional parameters that specify this flexible model: $\vec{\beta} = [\delta t, a_F, \eta_F, \varphi_S, \sigma_S, a_S, \eta_S, n, \kappa, k_F, f_{0F}, f_{MF}, k_S, f_{0S}, f_{MS}]$. The values of these parameters are listed in the table.

Training

The policy gradients are given by:

$$\begin{aligned} \nabla_{\vec{\omega}}\nu_F(\theta;\vec{\omega}) &= f'\left(\vec{\omega}_F^T\vec{g}(\theta)\right)\vec{g}(\theta) \\ \nabla_{\vec{\omega}}d_S(\theta;\vec{\omega}) &= f'\left(\vec{\omega}_S^T\vec{g}(\theta)\right)\vec{g}(\theta) \end{aligned} \qquad \text{(Equation 18)}$$

where $f'(x) = df/dx$ is the derivative of the sigmoidal function f. Together with Equations 13, 14, 15, and 16, these gradients can be used to compute the update to $\vec{\omega}_F$ and $\vec{\omega}_S$. Training thus acts to change the heading dependence in fixations and saccades by re-weighting the basis functions through changes in $\vec{\omega}_F$ and $\vec{\omega}_S$, as detailed in Algorithms 1–2.

In the main text, we used this flexible policy to determine how the average duration of fixations and direction of saccades should be structured as a function of heading in order to maintain a goal heading θ_G.

We used the following sensory response function that decays linearly with angular distance from the preferred heading:

$$R(\theta) = \frac{\pi}{10}\left(\frac{1}{2} - |\theta - \theta_G|\right) \qquad \text{(Equation 19)}$$

Algorithm 1. Learn flexible policy parameters $\vec{\omega}$ via policy-gradient method

input: parametrized policy $\pi(\dot{\theta}, \Delta t | \theta, \xi; \vec{\omega})$
define: total simulation time T_{tot}; fixed policy parameters $\vec{\beta}$
initialize: policy parameters $\vec{\omega} \in \mathbb{R}^d$; integrator $\xi = 0$; time $t = 0$; heading $\theta \in [0°, 360°]$

while $t < T_{tot}$ do
  sample action from policy
  $\dot{\theta}, \Delta t, \Delta\xi \sim \pi(\cdot | \theta, \xi; \vec{\omega})$
  observe sensory response
  $r \leftarrow R(\theta + \dot{\theta}\Delta t)$
  update policy parameters
  $\vec{\omega} \leftarrow \vec{\omega} + r\nabla_{\vec{\omega}} \log \pi(\dot{\theta}, \Delta t | \theta, \xi; \vec{\omega})$
  update heading, time, integrator
  $\theta \leftarrow \theta + \dot{\theta}\Delta t$
  $t \leftarrow t + \Delta t$
  $\xi \leftarrow \xi + \Delta\xi$
end while
return $\vec{\omega}$

Algorithm 2. Sample action $\dot{\theta}, \Delta t$ from flexible policy $\pi(\dot{\theta}, \Delta t | \theta, \xi; \vec{\omega})$

inputs: heading θ; integrator ξ; flexible policy parameters $\vec{\omega}$; fixed policy parameters $\vec{\beta}$; basis functions $\vec{g}(\theta)$

get current drift rate, directional bias
$\nu_F(\theta; \vec{\omega}_F) \leftarrow f(\vec{\omega}_F^T \vec{g}(\theta); k_F, f_{0,F}, f_{M,F})$
$d_S(\theta; \vec{\omega}_S) \leftarrow f(\vec{\omega}_S^T \vec{g}(\theta); k_S, f_{0,S}, f_{M,S})$
integrate drift signal
$\Delta\xi \sim \mathcal{N}(\nu_F(\theta; \vec{\omega}_F)\delta t, \eta_F^2 \delta t)$
if $\xi + \Delta\xi > a_F$ **then**
  saccade
  if rand$(\cdot) < d_S(\theta; \vec{\omega}_S)$ **then**
    $\dot{\theta} \sim +\text{logn}(\varphi_S, \sigma_S^2)$ (CW saccade)
  else
    $\dot{\theta} \sim -\text{logn}(\varphi_S, \sigma_S^2)$ (CCW saccade)
  end if
  $\Delta t \leftarrow t_S$
else
  fixate
  $\dot{\theta} \leftarrow 0$
  $\Delta t \leftarrow \delta t$
end if
return $\dot{\theta}, \Delta t, \Delta\xi$

Fixed-form policy tethered to a single goal heading

In Figure 6, we showed how a circuit-based implementation of a fixed-form, goal-heading-dependent policy (derived via the learning algorithm described in the previous section), when tethered to an internal representation of heading, could qualitatively capture the observed structure of fly behavior. Here, we illustrate a simplified form of this fixed-form policy, and we show how learning could act to shift the goal heading via a policy-gradient learning algorithm. In the following section, we show how this fixed-form policy could be implemented in a circuit model, and how learning could be implemented through Hebbian-like plasticity, rather than through the explicit computation of policy gradients.

Policy

In what follows, we will use θ_A to specify the angular orientation of the fly in arena coordinates, and we will distinguish this from the orientation θ_C of the compass heading bump. Before accounting for jumps in the compass bump, we assume that changes in the arena heading $\Delta\theta_A$ are accompanied by the same change in compass heading $\Delta\theta_C$, such that $\Delta\theta_A = \Delta\theta_C$ and $\dot{\theta}_A = \dot{\theta}_C = \dot{\theta}$ (in

the following section, we will explicitly account for the fact that in the fly heading circuit, the compass heading and the arena heading move in opposite directions during saccades). Bump jumps, which arise from symmetries in the visual environment, will further alter the compass heading θ_C relative to the arena heading θ_A.

The fixed-form policy specifies the heading dependence in the drift rate of fixations ν_F, and the directional bias of saccades d_S, relative to a goal heading $\theta_G \in [0°, 360°]$. Based on the results shown in Figure 3D, and given the prevalence of sinusoidal signals in the central complex, we define these dependencies to have the following functional forms:

$$\begin{aligned} d_S(\theta_C;\theta_G) &= -\frac{G_S}{2}\sin(\theta_C - \theta_G) + B_S + \frac{1}{2} \\ \nu_F(\theta_C;\theta_G) &= \frac{G_F}{2}(1 - \cos(\theta_C - \theta_G)) + B_F \end{aligned} \quad \text{(Equation 20)}$$

where B_S and G_S control the baseline direction and heading-dependence of saccades, and B_F and G_F similarly control the baseline duration and heading dependence of fixations. As before, the average duration of fixations at any given heading θ_C is given by $a_F/\nu_F(\theta_C;\theta_G)$, where a_F is the threshold of the drift diffusion process.

We showed in the main text that the heading representation can jump in symmetric scenes (Figure 2), and that the size of these jumps reflects the degree of symmetry in the scene. We consider the two-fold symmetric scene used in the main text, for which a bump jump results in an angular change of $\Delta\theta_J = \pm 180°$. We further assume that the probability of a jump varies non-uniformly with θ_C. We expect that the bump will have the lowest probability of jumping from locations through which the fly has frequently turned (as these are locations where the visual map from ring neurons onto compass neurons will have most quickly stabilized). Because the goal heading defines the angular location toward which the fly will turn, we define the probability of a jump $p_J(\theta_C;\theta_G)$ to be sinusoidal, with a value that depends on the angular distance from the goal heading:

$$p_J(\theta_C;\theta_G) = \frac{G_J}{2}(1 + \cos(\theta_C - \theta_G)) + B_J \quad \text{(Equation 21)}$$

where B_J and G_J control the baseline probability and heading-dependence of bump jumps. When G_J is 0, the bump has the same probability of jumping at every heading, determined by the value of B_J. For G_J larger than 0, the probability of a jump will vary non-uniformly with θ, and the minimum probability will be determined by B_J. The jump probability determines whether or not the heading will jump by $\Delta\theta_J$:

$$\theta_C = \begin{cases} \theta_C & \text{with probability } 1 - p_J(\theta_C;\theta_G) \\ \theta_C + \Delta\theta_J & \text{with probability } p_J(\theta_C;\theta_G) \end{cases} \quad \text{(Equation 22)}$$

From the perspective of a given arena heading θ_A, the effective drift rate and saccade probabilities are given by:

$$\begin{aligned} d_S^{\text{eff}}(\theta_A;\theta_G) &= (1 - p_J(\theta_A;\theta_G))d_S(\theta_A;\theta_G) + p_J(\theta_A;\theta_G)d_S(\theta_A + \Delta\theta_J;\theta_G) \\ \nu_F^{\text{eff}}(\theta_A;\theta_G) &= (1 - p_J(\theta_A;\theta_G))\nu_F(\theta_A;\theta_G) + p_J(\theta_A;\theta_G)\nu_F(\theta_A + \Delta\theta_J;\theta_G) \end{aligned} \quad \text{(Equation 23)}$$

As before, this policy depends implicitly on a set of fixed parameters $\vec{\beta}$, which now includes the additional parameters that specify the structure setpoint policy: $\vec{\beta} = [\delta t, a_F, \eta_F, \varphi_S, \sigma_S, a_S, \eta_S, G_S, G_F, G_J, B_S, B_F, B_J]$. The values of these parameters are listed in the table.

Training

The fixed-form policy can be trained using the same policy gradient algorithm discussed above (see Algorithms 3–3). We briefly discuss a simple algorithm for implementing this. We assume that the gain parameters $\{G_S, G_F, G_J\}$ and baseline parameters $\{B_S, B_F, B_J\}$ remain fixed during training, but that the goal heading θ_G is updated during training and only after a saccade. We further assume that bump jumps occur just after saccades, and that the goal heading is updated only after the bump has had an opportunity to jump. Thus, a single update consists of the following sequence of steps: (i) sample the duration of a fixation from an inverse Gaussian distribution, (ii) sample the direction and size of saccade, (iii) execute the change in heading in both arena and bump coordinates, (iv) flip a biased coin to determine whether the bump will jump; if the bump jumps, shift the bump location by 180°, and (v) update the location of the goal heading.

Because the goal heading is shifted only after a saccade (and not during the process of fixation), we can further simplify the process by directly sampling the fixation duration from an inverse Gaussian distribution; i.e., a fixation of total duration Δt is sampled directly from $P(\Delta t|\dot{\theta},\theta_C,F) = \text{IG}(a_F/\nu_F(\theta_C;\theta_G), a_F^2/\eta_F^2)$, rather than being generated in a timepoint-by-timepoint manner. We then sample a saccade of angular velocity $\dot{\theta} \sim \pi_S(\dot{\theta}|\theta_C;\theta_G)$ and fixed duration t_S (where $\pi_S(\dot{\theta}|\theta_C;\theta_G) = P(\dot{\theta}|\theta_C, S;\theta_G)$ is given in Equation 12). Finally, we update the location of the goal heading based on the gradient of the policy with respect to goal heading:

$$\Delta\theta_G\left(\dot{\theta}^*, \theta_A^*, \theta_C^*;\theta_G\right) = R\left(\theta_A + \dot{\theta}\Delta t_S\right)\frac{1}{\pi_S\left(\dot{\theta}\middle|\theta_C;\theta_G\right)}\left.\frac{\partial\pi_S\left(\dot{\theta}\middle|\theta_C;\theta_G\right)}{\partial\theta_G}\right|_{\dot{\theta}^*,\theta_A^*,\theta_C^*} \quad \text{(Equation 24)}$$

where $\Delta\theta_G(\dot{\theta}^*, \theta_A^*, \theta_C^*; \theta_G)$ specifies the change in the location of the goal heading that is produced by initiating a saccade with velocity $\dot{\theta}^*$ (and fixed duration t_S) from an arena heading θ_A^* and compass heading θ_C^*, given an initial goal heading of θ_G. The gradient is given by:

$$\begin{aligned}\frac{\partial \pi_S}{\partial \theta_G} &= \frac{\partial}{\partial \theta_G} P\left(\dot{\theta}\middle|\theta_C, S; \theta_G\right) \\ &= \mathrm{sgn}\left(\dot{\theta}\right) \mathrm{logn}\left(\left|\dot{\theta}\right|; \varphi_S, \sigma_S^2\right) \frac{\partial}{\partial \theta_G} d_S(\theta_C; \theta_G) \\ &= \frac{G_S}{2} \mathrm{sgn}\left(\dot{\theta}\right) \mathrm{logn}\left(\left|\dot{\theta}\right|; \varphi_S, \sigma_S^2\right) \cos(\theta_C - \theta_G)\end{aligned} \quad \text{(Equation 25)}$$

Circuit implementation of a fixed-form policy

In Figures 4, 5, and 6, we developed a circuit-based implementation of the fixed-form policy discussed in the previous section, and we used Hebbian learning to update the parameters of that model (note that this differs from the policy-gradient algorithms discussed above). We constructed this circuit model from populations of so-called columnar neurons that tile 360° of angular space. In what follows, we parametrize this space by θ, and we express all quantities as functions of θ. See additional resources section links to known anatomy for a discussion of the relationships between this model and anatomical and functional observations.

Policy

In the EB, the current heading θ_C is represented by a bump of activity maintained by a population of columnar compass neurons. We approximate this bump of activity using a von Mises function whose amplitude is normalized to 1:

$$r_C^{EB}(\theta, \theta_C) = \exp(\kappa \cos(\theta - \theta_C)) \,/\, 2\pi I_0(\kappa) \quad \text{(Equation 26)}$$

When the fly saccades by an angle $\Delta\theta_A$ in the arena, we assume that this is perfectly captured by the heading circuit, such that the current heading shifts by an equal but opposite angle of $\Delta\theta_C = -\Delta\theta_A$.

Flies are known to exhibit variability in the offset between the orientation of the heading bump and the orientation of the visual scene; we assume here that this offset is zero. In visual scenes without repeating patterns, plasticity between ring neurons and compass neurons ensures that this offset remains stable over time by reinforcing a relationship between visual features in the scenes (as conveyed through the ring neuron receptive fields) and the location of the heading bump. However, in scenes with repeating patterns, ring neurons with a given receptive field will respond similarly when the fly is oriented toward different symmetric views of the same scene. This will result in the strengthening of different sets of ring-to-compass-neuron weights that correspond to the same visual patterns, and will lead to jumps in the heading representation (Figure 2D). We incorporate these dynamics through a set of heading-dependent compass weights $\vec{\omega}_C = \omega_C(\theta_A, \theta_B)$ that capture the net inhibition from each ring neuron whose receptive field is positioned at θ_A onto each compass neuron with preferred heading θ_B, and thereby determine the relative stability of different headings that correspond to symmetric views of the visual scene. For the visual scenes considered here, there is a two-fold symmetry, such that the fly sees identical views of the same scene at the two orientations θ and $\theta + \Delta\theta_J$, where $\Delta\theta_J = 180°$. These identical views will both activate the same two subsets of rings neurons; we assume here the scene fully activates rings neurons whose receptive fields are positioned at θ and $\theta + \Delta\theta_J$, and partially activates ring neurons whose receptive fields are positioned at $\theta \pm \delta\theta$ and $\theta + \Delta\theta_J \pm \delta\theta$, where $\delta\theta$ is the spacing between preferred headings. We take the net weight profile summed across these active ring neurons, $\omega_{\mathrm{net}}(\theta) = \sum_{\mathrm{act}} \omega_C(\theta_{A,\mathrm{act}}, \theta)$, and we compare it between the two symmetric compass headings, θ_C and $\theta_C + \Delta\theta_J$. The higher the net weight at the current compass heading, the stronger the inhibition from the ring neurons, and the more likely the bump will jump by $\Delta\theta_J$ to the symmetric compass heading. We define this jump probability to be:

$$p_J\left(\theta_C; \vec{\omega}_C\right) = \frac{\omega_{\mathrm{net}}(\theta_C)}{\omega_{\mathrm{net}}(\theta_C) + \omega_{\mathrm{net}}(\theta_C + \Delta\theta_J)} \quad \text{(Equation 27)}$$

From the EB, the heading representation travels through the protocerebral bridge to the fan-shaped body (FB), where the profile of compass activity takes on a sinusoidal shape:

$$r_C^{FB}(\theta, \theta_C) = \frac{1}{2}(\cos(\theta - \theta_C) + 1) \quad \text{(Equation 28)}$$

We assume that information about the fly's current compass heading is combined with information about the goal heading in the FB and then used to drive premotor activity in the lateral accessory lobe (LAL). Specifically, we assume that the information about the goal heading is stored in a set of synaptic weights $\vec{\omega}_G = \omega_G(\theta)$ from tangential motor-state neurons onto columnar goal neurons. Here, we assume that the motor state neurons are active (with a constant activity of one; i.e., $r_M(\theta) = 1 \ \forall_\theta$) whenever the fly is flying. Thus, the activity profile $r_G(\theta)$ of the goal neurons gives a direct readout of the goal weights:

$$\begin{aligned} r_G\left(\theta; \vec{\omega}_G\right) &= r_M(\theta)\omega_G(\theta) \\ &= \omega_G(\theta) \end{aligned} \quad \text{(Equation 29)}$$

As we will detail below, the set of weights $\vec{\omega}_G$ fully determines the properties of the goal heading.

Finally, we consider populations of output neurons that receive goal activity $r_G(\theta)$ and phase-shifted heading activity $r_C(\theta, \theta_C + \vartheta)$ as inputs, and whose summed output depends on the overlap between current and goal heading through a multiplicative operation:

$$r_O\left(\theta_C, \vartheta; \vec{\omega}_G\right) = \frac{\sum_\theta r_C^{FB}(\theta, \theta_C+\vartheta) r_G\left(\theta; \vec{\omega}_G\right)}{\sum_\theta r_C^{FB}(\theta, \theta_C+\vartheta) r_C^{FB}(\theta, \theta_C+\vartheta)} + B_O \quad \text{(Equation 30)}$$

where ϑ is a phase shift, and B_O is a baseline shift (described below). The form of the output activity in Equation 30 ensures that this output activity will be structured sinusoidally as a function of the fly's current compass heading θ_C relative to the circular mean of the goal weights. To see this, note that the numerator of Equation 30 can be written as:

$$\begin{aligned}
\text{num}\left[r_O\left(\theta_C, \vartheta; \vec{\omega}_G\right)\right] &= \sum_\theta \frac{1}{2}(1 + \cos(\theta - (\theta_C + \vartheta)))\omega_G(\theta) \\
&= \frac{1}{2}\sum_\theta \omega_G(\theta) + \frac{1}{2}\left(\frac{e^{-i(\theta_C+\vartheta)}}{2}\sum_\theta e^{i\theta}\omega_G(\theta) + \frac{e^{i(\theta_C+\vartheta)}}{2}\sum_\theta e^{-i\theta}\omega_G(\theta)\right) \\
&= \frac{|\omega_G|}{2} + \frac{r_G}{2}\left(\frac{e^{-i(\theta_C+\vartheta-\theta_G)} + e^{i(\theta_C+\vartheta-\theta_G)}}{2}\right) \\
&= \frac{|\omega_G|}{2} + \frac{r_G}{2}\cos(\theta_C + \vartheta - \theta_G)
\end{aligned} \quad \text{(Equation 31)}$$

where $\sum_\theta e^{i\theta}\omega_G(\theta)d\theta \equiv r_G e^{i\theta_G}$ specifies the modulus r_G and angle θ_G of the circular mean of the goal weights, and $|\omega_G| = \sum_\theta \omega_G(\theta)$ specifies the total strength of the goal weights. As we will next show, saccades will drive the compass heading toward θ_G, and fixations will be maintained longer at θ_G, and thus we define θ_G to be the goal heading. The denominator of Equation 30 is fixed; we will denote this as $D = \sum_\theta r_C^{FB}(\theta, \theta_C + \vartheta) r_C^{FB}(\theta, \theta_C + \vartheta)$.

The dynamic range of the output activity determines how strongly the goal drives behavior; the larger the range, the bigger the differential between the behavior at versus away from the goal location. In such cases with a large differential, we describe the behavior as "highly structured". As can be seen in Equation 31, this range will be maximized when the modulus of the circular mean, r_G, is maximized. Because we constrain the heading and goal weights to lie in the range [0,1] (more on this below), the weight profile that maximizes r_G is a square wave in which $N/2$ consecutive weights take a value of 0, and the remaining $N/2$ consecutive weights take a value of 1. Below, we describe a learning rule that will drive the weights toward a sinusoidal profile with a single peak, which approximates this square wave profile. Thus, within the constraints of this learning rule, the more strongly sinusoidal the weight profile, the more structured the behavior.

We construct three different output populations, denoted left ('L'), fixation ('F'), and right ('R'), that differ in the phase shift of their heading-tuned inputs[24]:

$$\begin{aligned}
r_L\left(\theta_C; \vec{\omega}_G\right) &= r_O\left(\theta_C, \vartheta = +90; \vec{\omega}_G\right) = \frac{|\omega_G|}{2D} + \frac{r_G}{2D}\cos(\theta_C + 90 - \theta_G) \\
r_F\left(\theta_C; \vec{\omega}_G\right) &= r_O\left(\theta_C, \vartheta = 180; \vec{\omega}_G\right) = \frac{|\omega_G|}{2D} + \frac{r_G}{2D}\cos(\theta_C + 180 - \theta_G) + B_C \\
r_R\left(\theta_C; \vec{\omega}_G\right) &= r_O\left(\theta_C, \vartheta = -90; \vec{\omega}_G\right) = \frac{|\omega_G|}{2D} + \frac{r_G}{2D}\cos(\theta_C - 90 - \theta_G)
\end{aligned} \quad \text{(Equation 32)}$$

where we chose $B_L = B_R = 0$, $B_C = \nu_{\max} - |\omega_G|/D$, and $\nu_{\max} = 1.1$; as described below, this ensures that gain of fixations will increase with increasing $|\omega_G|$. We further assume, based on known projection patterns,[24] that the populations of left and right output neurons project unilaterally to descending neurons that control leftward and rightward saccades, respectively, and that the population of center output neurons projects bilaterally to both sets of descending neurons. We assume that the activity of the right and left output neurons thus determines the average directionality of saccades for a given heading θ:

$$\begin{aligned}
d_S\left(\theta_C; \vec{\omega}_G\right) &= \frac{1}{2}\left(1 + r_R\left(\theta_C; \vec{\omega}_G\right) - r_L\left(\theta_C; \vec{\omega}_G\right)\right) \\
&= \frac{1}{2} + \frac{r_G}{D}\cos(\theta_C - 90 - \theta_G)
\end{aligned} \quad \text{(Equation 33)}$$

Note that this function will be largest, and will thus drive the highest probability of clockwise saccades (and counterclockwise bump rotations), when the heading is 90° to the right of the goal heading.

We similarly assume that the activity of the center output neurons determines the drift rate (and thereby the average duration) of fixations:

$$\begin{aligned}\nu_F\left(\theta_C;\overrightarrow{\omega}_G\right) &= r_F\left(\theta_C;\overrightarrow{\omega}_G\right)\\ &= \nu_{\max} - \frac{|\omega_G|}{2D} + \frac{r_G}{2D}\cos(\theta_C + 180 - \theta_G)\end{aligned} \quad \text{(Equation 34)}$$

$$\langle \Delta t_F\left(\theta_C;\overrightarrow{\omega}_G\right)\rangle = \frac{1}{\nu_F\left(\theta_C;\overrightarrow{\omega}_G\right)} \quad \text{(Equation 35)}$$

Thus, the baseline values of saccade and fixation properties are given by:

$$\begin{aligned}\min[d_S] &= \frac{1}{2} - \frac{r_G}{D}\\ \min[\nu_F] &= \frac{2D\nu_{\max} - |\omega_G| - r_G}{2D}\\ \min[\langle \Delta t_F\rangle] &= \frac{2D}{2D\nu_{\max} - |\omega_G| + r_G}\end{aligned} \quad \text{(Equation 36)}$$

and the gain of saccade and fixation properties are given by:

$$\begin{aligned}\text{gain}[d_S] &= \frac{2r_G}{D}\\ \text{gain}[\nu_F] &= \frac{r_G}{D}\\ \text{gain}[\langle \Delta t_F\rangle] &= \frac{r_G/D}{\nu_{\max}^2 + |\omega_G|^2 \Big/ \left(4D^2\right) - r_G^2 \Big/ \left(4D^2\right) - \nu_{\max}|\omega_G| \Big/ D}\end{aligned} \quad \text{(Equation 37)}$$

As can be seen from Equations 33, 34, and 35, the modulus r_G of the circular mean of the goal weights determines the gain of both the fixational drift rate and the saccade directionality; the larger r_G, the larger the directional bias when the heading bump is to the right or left of the goal heading, the longer the fixations at the goal heading, and the shorter the fixations away from the goal heading. In this way, the multiplicative operation performed by the output neuron populations (and specified by Equation 31) guarantees that the fly's internal policy will remain structured as a function of the current heading relative to the goal heading, regardless of their specific values.

Training

We assume that there is plasticity in both $\overrightarrow{\omega}_C$ and $\overrightarrow{\omega}_G$ that is mediated by the activity of the heading bump in the EB and FB ($r_C^{EB}(\theta,\theta_C)$ and $r_C^{FB}(\theta,\theta_C)$, respectively):

$$\begin{aligned}\Delta\omega_C\left(\theta,\theta_C;\overrightarrow{\omega}_C\right) &= \alpha_C\Delta_C v^2\\ \Delta\omega_G\left(\theta,\theta_C;\overrightarrow{\omega}_G\right) &= \alpha_G\Delta_G\end{aligned} \quad \text{(Equation 38)}$$

where

$$\begin{aligned}\Delta_C &= \left[r_C^{EB}(\theta,\theta_C) - \omega_C(\theta)\right]_+\Theta(1-\omega_C) - \left[\omega_C(\theta) - r_C^{EB}(\theta,\theta_C)\right]_+\Theta(\omega_C)\\ \Delta_G &= \begin{cases} +\left[r_C^{FB}(\theta,\theta_C) - \omega_G(\theta)\right]_+\Theta(1-\omega_G) - \left[\omega_G(\theta) - r_C^{FB}(\theta,\theta_C)\right]_+\Theta(\omega_G) & R(\theta_A) > 0\\ -\left[r_C^{FB}(\theta,\theta_C) - \omega_G(\theta)\right]_+\Theta(\omega_G) + \left[\omega_G(\theta) - r_C^{FB}(\theta,\theta_C)\right]_+\Theta(1-\omega_G) & R(\theta_A) < 0\\ 0 & R(\theta_A) = 0\end{cases}\end{aligned} \quad \text{(Equation 39)}$$

Here, $[\cdot]_+$ denotes rectification, and $\Theta(\cdot)$ is the heaviside function.

The first of these plasticity rules is similar to that used in Kim et al.[44] in that the change in weights is proportional to the co-activity between ring neurons (whose activity here is implicilty conveyed through the compass weights) and compass neurons (whose activity here is assumed to have a fixed profile $r_C^{EB}(\theta,\theta_C)$), and is modulated by fly's velocity v. In simulations, we use the size of the saccade, $\Delta\theta_S$, as a proxy for this velocity. We assume that the compass weights are only updated only during saccades, and that the goal weights are updated only during fixations. In practice, we partition saccades into angular increments $\delta\theta$ (i.e., based on the spacing between preferred headings), and we partition fixations into time increments of 100 ms. We then iteratively update weights at each angle/time increment. The second plasticity rule differs from the first in that it additionally

incorporates the valence, $R(\theta_A)$, of the current arena heading θ_A. We assume that this valence is carried by tangential neuromodulatory neurons that innervate the FB and themselves receive input from heading-tuned neurons.[24] We assume this valence takes the following form:

$$R(\theta_A) = \begin{cases} +1 & \theta_A \in \text{safe} \\ -1 & \theta_A \in \text{danger} \end{cases} \quad \text{(Equation 40)}$$

Algorithms 3–4 detail how this circuit model is implemented and updated through training.

Algorithm 3. Learn heading and goal weights, $\omega_C(\theta)$ and $\omega_G(\theta)$

input: parameterized policy $\pi(\Delta\theta, \Delta t | \theta; \vec{\omega}_C, \vec{\omega}_G)$
define: total simulation time T_{tot}, fixed policy parameters $\vec{\beta}$, learning rates α_C, α_G, heading resolution $\delta\theta$
initialize: weights $\vec{\omega}_C$ and $\vec{\omega}_G$; compass heading $\theta_C \in [0°, 360°]$; arena heading $\theta_A = \theta_C$; time $t = 0$, $\Delta t = 0$;

while $t < T_{tot}$ **do**
  sample action from policy
  $[\Delta\theta_S, \Delta\theta_F], [\Delta t_S, \Delta t_F] \sim \pi(\cdot | \theta_C; \vec{\omega}_C, \vec{\omega}_G)$
  update arena heading, compass heading, and compass weights during a saccade
  $t \leftarrow t + \Delta t_S$
  $\Delta\theta = 0$
  while $\Delta\theta < \Delta\theta_S$ **do**
    $\omega_C(\theta) \leftarrow \omega_C(\theta) + \alpha_C \Delta_C \Delta\theta_S^2$
    $\theta_C \leftarrow \theta_C + \delta\theta$
    $\theta_A \leftarrow \theta_A - \delta\theta$
    $\Delta\theta = \Delta\theta + \delta\theta$
  end while
  determine whether bump will jump
  if rand$(\cdot) < p_{\text{jump}}(\theta_C; \vec{\omega}_C)$ **then**
    $\theta_C \leftarrow \theta_C + \Delta\theta_C$
  end if
  observe sensory response
  $r \leftarrow R(\theta_A)$
  update goal weights during fixation
  while $\Delta t < \Delta t_F$ **do**
    $\omega_G(\theta) \leftarrow \omega_G(\theta) + \alpha_G \Delta_G$
    $\Delta t \leftarrow \Delta t + 0.1$
  end while
  $\Delta t = 0$
end while
return $\vec{\omega}_C$, $\vec{\omega}_G$

Algorithm 4. Sample action sequence $[\Delta\theta_S, \Delta\theta_F]$, $[\Delta t_S, \Delta t_F]$ from setpoint policy $\pi(\Delta\theta, \Delta t | \theta; \vec{\omega}_C, \vec{\omega}_G)$

inputs: heading θ_C; weights $\vec{\omega}_C$, $\vec{\omega}_G$

get current drift rate and directional bias
$\nu_F(\theta_C; \vec{\omega}_G) \leftarrow r_F(\theta_C; \vec{\omega}_G) + B_F$
$d_S(\theta_C; \vec{\omega}_G) \leftarrow (1 + r_R(\theta_C; \vec{\omega}_G) - r_L(\theta_C, \vec{\omega}_G))/2$
saccade
if rand$(\cdot) < d_S(\theta_C; \vec{\omega}_G)$ **then**
  $\Delta\theta_S \sim +\text{logn}(\varphi_S, \sigma_S^2)$ (CW)
else
  $\Delta\theta_S \sim -\text{logn}(\varphi_S, \sigma_S^2)$ (CCW)
end if
$\Delta t_S \leftarrow t_S$
fixate
$\Delta\theta_F \leftarrow 0$
$\Delta t_F \leftarrow 1/\nu_F(\theta_C; \vec{\omega}_G)$
return $[\Delta\theta_S, \Delta\theta_F]$, $[\Delta t_S, \Delta t_F]$