

# Multimodal cue integration and learning in a neural representation of head direction

Received: 25 December 2022

Accepted: 17 October 2024

Published online: 23 July 2025

Melanie A. Basnak, Anna Kutschireiter , Tatsuo S. Okubo , Albert Chen , Pavel Gorelik, Jan Drugowitsch  & Rachel I. Wilson  ✉

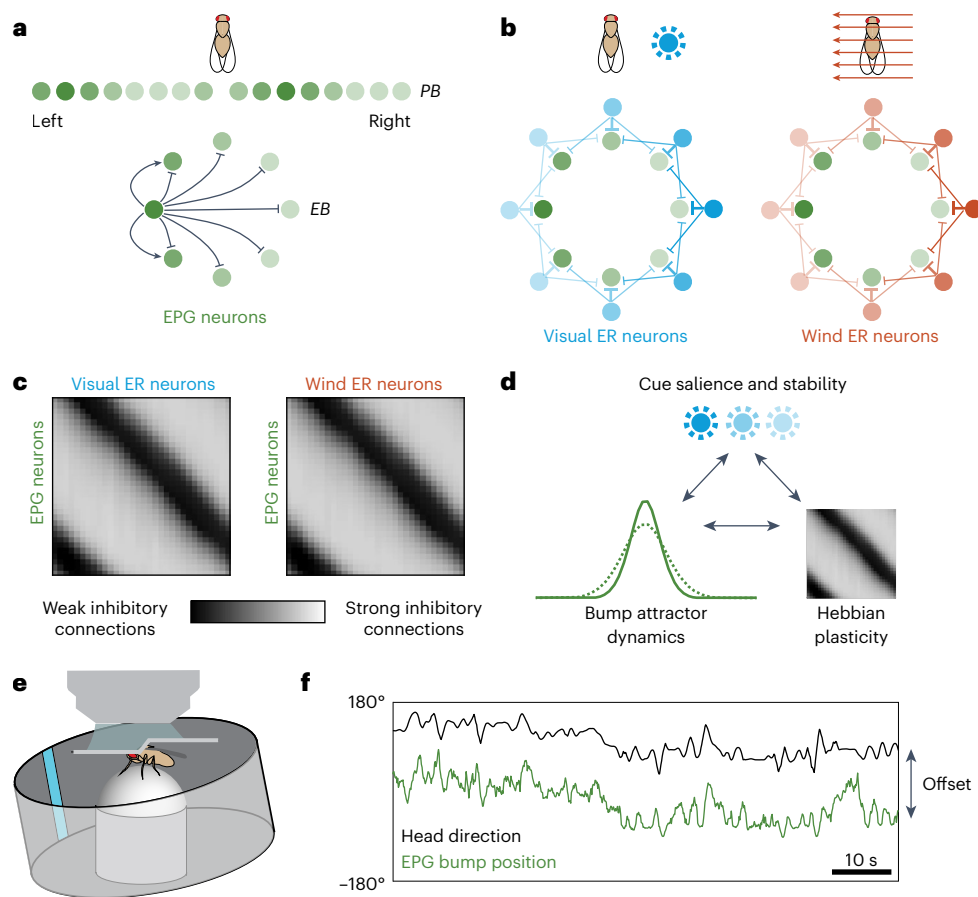
Navigation requires us to take account of multiple spatial cues with varying levels of informativeness and learn their spatial relationships. Here we investigate this process in the *Drosophila* head direction system, which functions as a ring attractor and a topographic map of head direction. Using population calcium imaging and multimodal virtual reality environments, we show that increasing cue informativeness improves encoding accuracy and produces a narrower and higher bump of activity. When cues conflict, the more informative cue exerts more weight. A familiar cue is weighted more heavily and used to guide the remapping of a less familiar cue. When a cue is less informative, it is remapped more readily in response to cue conflict. All these results can be explained by an attractor model with plastic sensory synapses. Our findings provide a mechanistic explanation for how the brain assembles spatial representations through inference and learning.

When we enter a new environment, we encounter a host of cues that might be useful for guiding navigation. Some cues are more useful than others because they are more salient or easily located; for example, the moon is more visible than a star. At the same time, the usefulness of a spatial cue also depends on its stability<sup>1</sup>. For example, a strong wind is salient because it can easily inform us about the direction we are facing, but only if the wind is blowing from a stable direction. Similarly, a faraway mountain and a nearby tree may be equally salient as visual objects but only the mountain will have a stable position on the horizon and, thus, it is much more informative about the direction we are facing. For this reason, we should assess the stability of each external cue by monitoring its position over time, relative to our own self-motion cues. As we acquire more familiarity with an external cue, we should logically ascribe it more weight (relative to self-motion cues), as long as the external cue appears stable<sup>1,2</sup>. Thus, a cue must be both stable and familiar to be highly informative for navigation.

Not surprisingly, behavioral studies have shown that navigating animals generally ascribe more weight to external cues that appear to be more stable and familiar, just as they ascribe more weight to cues that are more salient<sup>3–14</sup>. Moreover, neurophysiological studies in rats and mice have shown that salient or stable and familiar cues exert the strongest influence over the neurons in the brain's navigation centers<sup>15–17</sup>. How does this work mechanistically? The brain's navigation centers are thought to be organized around attractor networks (that is,

networks with multiple stable states)<sup>18</sup>. Self-motion cues drive transitions between different stable states, creating a working memory of the organism's position on a map or its orientation in space. Specifically, the attractors that correspond to the sense of direction are thought to be ring attractors, meaning that their stable states form a closed circle in network state space<sup>19,20</sup>. Self-movement signals during head rotations would push the network state around this circle. It has been proposed that the sensory weights onto these ring attractors might be dynamically adjusted according to a Hebbian learning rule<sup>20–28</sup>. This would automatically allocate more weight (stronger synaptic connections) to environmental cues that are more salient, stable and familiar. Once stable environmental cues are familiar and, thus, well learned, they should improve the ring attractor network's ability to accurately track head direction (HD) beyond what it can achieve on the basis of self-movement signals alone<sup>19,20</sup>. Although this model was originally proposed to explain HD cells, it has also been extended to explain grid cells<sup>29–31</sup>. However, all of these models have been largely untested at a mechanistic level.

Recently, the *Drosophila* HD network has emerged as a useful system for testing these ideas. Genetic experiments have demonstrated that *Drosophila* HD cells are essential for navigation behavior<sup>32,33</sup>. All the HD cells in this network can be imaged simultaneously as a head-fixed fly navigates in a virtual reality environment<sup>33,34</sup>. Moreover, this network's anatomical connectivity is known in detail from connectome



**Fig. 1 | The *Drosophila* HD system. a**, HD cells (EPG neurons) form a ring attractor network in the EB. Their axons project to the PB, where they form two linearized topographic maps of HD. **b**, The position of the EPG activity bump is influenced by ER neurons that encode the positions of visual HD cues or the direction of the wind. ER neurons are inhibitory and the most active ER neurons push the bump to the location where their inhibitory output is minimal. ER → EPG connections are anatomically all-to-all but their weights are shaped by Hebbian plasticity at ER → EPG synapses, such that each ER neuron generally makes functional synapses onto only a subset of EPG neurons. **c**, Schematic ER → EPG weights. Given a single visual cue and a steady wind direction, associative LTD is predicted to produce a diagonal notch of weak connections in each weight matrix. EPG neurons are sorted by their preferred HD. ER neurons are sorted

by their preferred cue position. If the two cues are aligned in the simulated environment, Hebbian plasticity should align the notches. **d**, We hypothesize that cue salience and stability affect bump attractor dynamics and learning. **e**, We image EPG neurons in head-fixed flies walking on a spherical treadmill. As the fly turns on the spherical treadmill, the virtual environment rotates around the fly in the expected direction. Here, the environment contains a bright vertical stripe that serves as an HD cue. **f**, The bump of EPG activity tracks the fly's fictive HD in a virtual reality environment, with a relatively constant angular offset. Bump position rotates clockwise in the EB (imaged from the posterior side of the head) as HD rotates counterclockwise; therefore, to account for this directionality, we always plot (−HD) to make it easier to visualize the correspondence between bump position and HD.

data<sup>35</sup>. This network functions as a ring attractor, which is also a topographic map<sup>36</sup> (a ‘bump attractor’); it exhibits a persistent bump of activity, whose position stores a working memory of the fly's current orientation (Fig. 1a). The bump's position changes smoothly as the fly rotates, reflecting the influence of self-motion cues<sup>37,38</sup> and external sensory inputs<sup>34,39</sup> (Fig. 1b). The pattern of connection weights from sensory cells onto HD cells can change during spatial learning according to a Hebbian learning rule, allowing this system to learn the pattern of visual cues in the current environment<sup>22–24</sup> (Fig. 1c).

Here, we ask how the function of this network depends on cue salience and familiarity (Fig. 1d). We find that cue salience and familiarity alter the width and amplitude of the bump while also driving rapid spatial learning that continuously updates the properties of the network, such that salient and familiar cues are accorded more weight. Mechanistically, our results can be explained by a ring attractor model with a high rate of synaptic modification at sensory synapses onto HD cells. Conceptually, our findings show how continuous synaptic plasticity allows ongoing spatial learning and inference in a dynamic environment, albeit at the cost of reducing the stability of the system's representational coordinate frame. Thus, our results highlight the

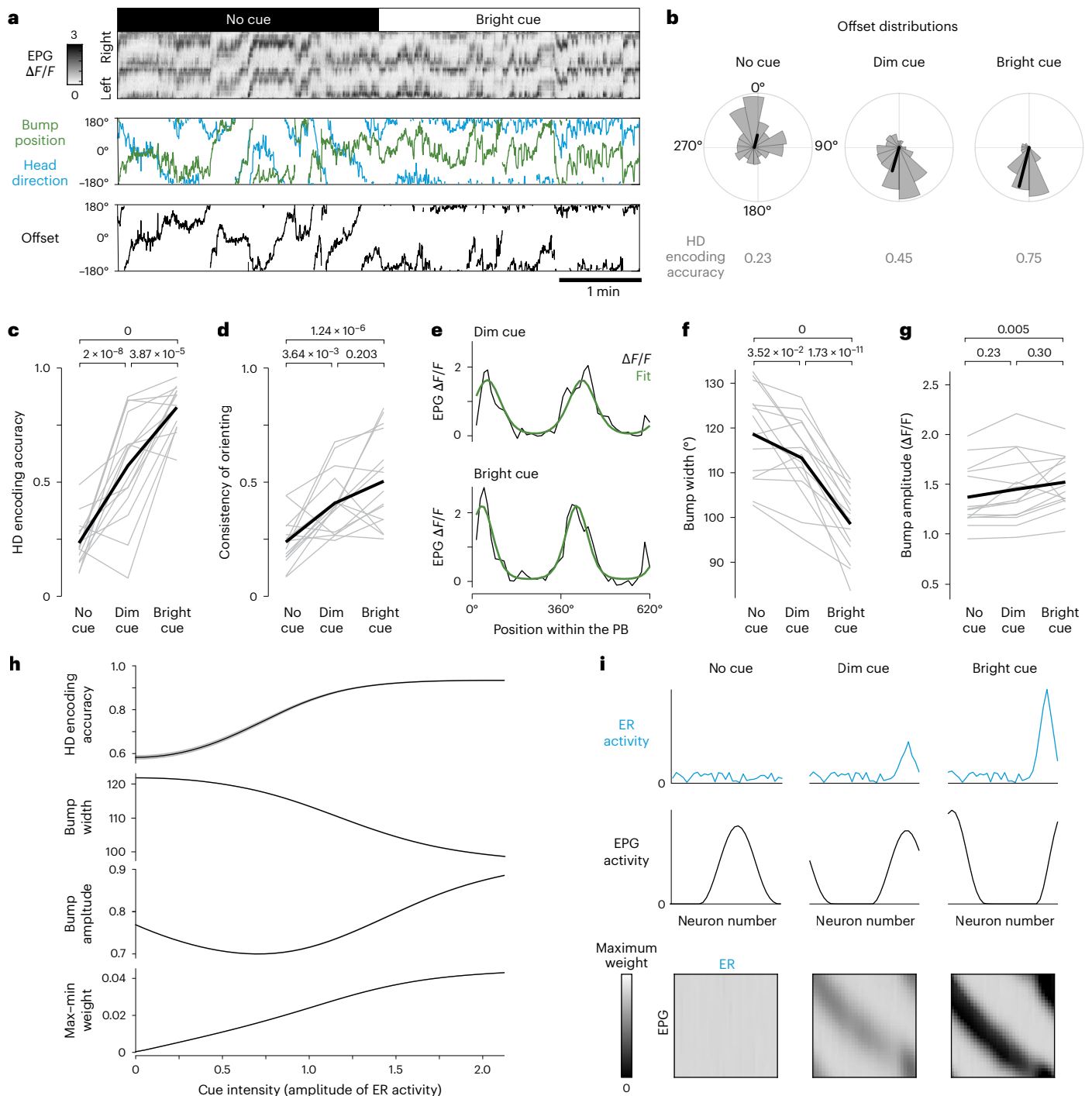
fundamental tradeoff between stability and flexibility in the brain's navigational centers.

## Results

To study cue integration and learning in the *Drosophila* HD system, we expressed jGCaMP7f (ref. 40) in HD cells (equivalence potential gradient (EPG) neurons) under Gal4–UAS control<sup>41</sup> and we imaged the EPG ensemble using a two-photon microscope as the fly walked freely on a spherical treadmill (Fig. 1e). By measuring the rotational velocity of the sphere, we can infer the fly's intended rotational velocity. If we then rotate a direction cue around the fly in closed loop with the ball's rotation, the EPG ensemble can track the fly's fictive HD (Fig. 1f). Note that the head and body are rigidly coupled in our experiments; thus, HD is always equal to heading. The fly rotates its fictive HD and heading by maneuvering on the spherical treadmill.

### Increasing cue intensity changes the bump profile

First, we examined how cue salience affects bump attractor dynamics. In these experiments, we used intensity as a proxy for salience; we switched among a bright cue, a dim cue and no cue in randomized



**Fig. 2 | Increasing cue intensity changes the bump profile.** This figure shows data for 15 flies. **a**, EPG  $\Delta F/F$  in the PB is used to infer the position of the EPG bump in the EB (green). When a visual cue is present, the bump tracks HD (blue). The offset between HD and bump is fairly constant when a bright visual HD cue is available. **b**, Offset distributions for the fly in **a**. Black lines are vector averages, with line length denoting HD encoding accuracy. Polar histograms are shown on the same scale. **c**, HD encoding accuracy in each condition. In **c**, **d**, **f**, **g**, single flies are shown in gray, mean values are shown in black and *P* values are shown (linear mixed-effects models with Tukey comparisons and Bonferroni corrections).

**d**, Consistency of behavioral orienting in each condition. **e**,  $\Delta F/F$  and fit (Extended Data Fig. 1) at two time points for the same fly. **f**, Bump width in each condition. **g**, Bump amplitude in each condition. **h**, Model of the effect of increasing cue intensity on HD encoding accuracy, EPG bump width, EPG bump amplitude and the range of ER → EPG weights (max-min). Mean of 100 simulation runs  $\pm$  s.e.m. **i**, Model of the spatial profile of ER and EPG population activity at one HD for each condition. Bottom, ER → EPG weights at the end of each block. The maximum weight represents maximum inhibition. The bright cue produces a deeper notch in the weight matrix.

interleaved 200-s blocks. For each cue intensity, we measured the accuracy of HD encoding (Fig. 2a). A perfectly accurate HD system should have a constant offset between HD and bump position. We, therefore, measured the circular variance in the offset over the duration of each block and we defined ‘HD encoding accuracy’ as  $1 - \text{circular variance}$

(Fig. 2b). We cannot directly measure the informativeness of a spatial cue by measuring its physical properties<sup>14</sup> but we can use HD encoding accuracy as an operational measure of a cue’s informativeness.

In pilot experiments, we lowered the intensity of the dim cue until HD encoding accuracy was only slightly better than no cue at all. We

then systematically compared HD encoding accuracy for all three cue conditions and we confirmed that HD encoding accuracy progressively improves with higher cue intensity (Fig. 2c); this demonstrates that the bright cue actually conveys more information. This makes sense because a bright cue should have a higher signal-to-noise ratio. Moreover, with higher cue intensity, we found that flies also tended to orient in a more consistent direction in virtual space (Fig. 2d). If the brain's internal estimate of HD is more accurate, it is logical that this should enable a more consistent behavioral orientation because flies use their HD system to orient toward an internal goal direction<sup>32,33,42</sup>.

Notably, when we increased cue intensity, we also found changes in the bump profile. Specifically, bump width decreased in a graded manner as visual cue intensity increased (Fig. 2e,f), mirroring the graded increase in HD encoding accuracy. Meanwhile, bump amplitude increased (Fig. 2g), although this effect was only clear for the highest cue intensity.

To better understand why cue intensity should affect the bump profile, we modeled this network as a ring attractor with plastic sensory inputs. In this model, each EPG neuron excites its neighbors while also driving global inhibition<sup>35,36</sup> (Fig. 1a). EPG neurons receive inhibitory input from ER neurons<sup>23</sup>, whose receptive fields tile visual space<sup>43</sup> (Fig. 1b). In agreement with anatomical data<sup>35</sup>, ER → EPG connections are all-to-all, with weights governed by a Hebbian learning rule<sup>22–24</sup> that weakens inhibition between coactive ER–EPG pairs through associative long-term depression (LTD). At the same time, this learning rule also strengthens inhibition onto EPG neurons that are active without ER input through nonassociative long-term potentiation (LTP), which depends on postsynaptic activity alone. EPG neurons also receive noisy self-motion signals that tend to push the EPG bump in the correct direction during turning maneuvers<sup>37,38</sup>. For each simulation iteration, we generated a random sequence of turning maneuvers that specify HD, ER input and self-motion input.

In this model, increasing the intensity of a visual cue increases the accuracy of HD encoding (Fig. 2h). It also decreases bump width (Fig. 2h). This is because ER neurons are inhibitory and increasing the intensity of a visual cue recruits more inhibitory drive to the network. Meanwhile, in this model, increasing visual cue intensity produces nonmonotonic changes in bump amplitude, reflecting two competing effects. On the one hand, increasing inhibitory drive pushes bump amplitude down. On the other hand, increasing presynaptic activity promotes associative LTD, producing a deeper notch in the pattern of ER → EPG weights (Fig. 2i), leading to stronger disinhibition of the most active EPG neurons; this pushes bump amplitude up. If the cue is very bright, the latter effect wins, producing a net increase in bump amplitude (Fig. 2h). In essence, the pattern of ER → EPG weights resembles the negative image of the cue and the intensity of that image reflects the intensity of the cue.

To summarize, we find that increases in cue brightness increase HD encoding accuracy and narrow the bump. Very bright cues also increase bump amplitude. All these changes in bump profile can be explained by the interaction of inhibitory sensory input and a Hebbian learning rule.

Thus far, we focused on average trends across flies but it is also instructive to examine individual differences in bump dynamics. In particular, we noticed that the same virtual reality environment could produce high HD encoding accuracy in some individuals (Fig. 3a) but lower accuracy in other individuals (Fig. 3b). Moreover, these variations were correlated with individual differences in bump width (Fig. 3c). We can recapitulate these results in the model by generating individual variations in the overall level of visually evoked activity in ER neurons. This produces variations in HD encoding accuracy and correlated changes in bump width (Fig. 3d). Bump amplitude variations are not so well correlated with HD encoding accuracy, both in our data (Fig. 3e) and in the model (Fig. 3f). In the model, this is because an increase in ER activity can produce opposing effects on bump amplitude (Fig. 2h).

In short, our results show that different individuals can experience the same cue as conveying more or less information. This produces correlated variations across individuals in bump width and HD encoding accuracy. Our model suggests that these individual variations arise with differences in the intensity of sensory input to the HD system, which might arise from individual differences in how flies process the visual cue.

### More informative cues are accorded more weight

Next, to investigate how different cues are integrated, we introduced wind into our virtual reality environments. We delivered wind through a tube that we rotated around the fly whenever the fly turned on the spherical treadmill, such that the environmental wind direction appeared constant from the fly's perspective. At the outset of each experiment, we presented the visual cue and the wind cue alone (Fig. 4a) and we confirmed that they produced similar HD encoding accuracy on average (Fig. 4b). However, in some individuals, HD encoding accuracy was higher with the visual cue, whereas, in other individuals, HD encoding accuracy was higher with the wind cue. In general, the cue that generated better HD encoding accuracy was the cue that produced the narrower and higher-amplitude bump (Fig. 4c). Thus, one cue was often experienced as conveying more information, although the two cues were equally informative on average.

After we combined the cues, we shifted one or the other every few minutes to create a conflict between them and to see which cue carried more weight. We saw that the bump sometimes followed the shifted cue while, in other instances, it did not move (Fig. 4d). Across all trials and all individuals, visual shifts and wind shifts had a similar influence on the bump (Extended Data Fig. 3). Nonetheless, some individuals systematically gave more weight to the visual cue, meaning that the bump followed visual shifts more than wind shifts. In general, these were the individuals where the visual cue produced higher HD encoding accuracy when it was presented alone (Fig. 4e). Conversely, other individuals gave more weight to the wind; these were the individuals where the wind produced higher HD encoding accuracy when presented alone (Fig. 4e). In short, the cue that produced better HD encoding was generally accorded more weight in cases of cue conflict.

Our model can reproduce our results if we have two populations of ER neurons and we vary the ER amplitude of one or the other population. When each cue is presented individually, this produces correlated individual variations in HD encoding accuracy and bump width, as well as weak effects on bump amplitude (Figs. 2h and 4f). When the cues are presented together and one is shifted, we find that the relative HD encoding accuracy of the two cues is a good predictor of the bump preference index (Fig. 4g).

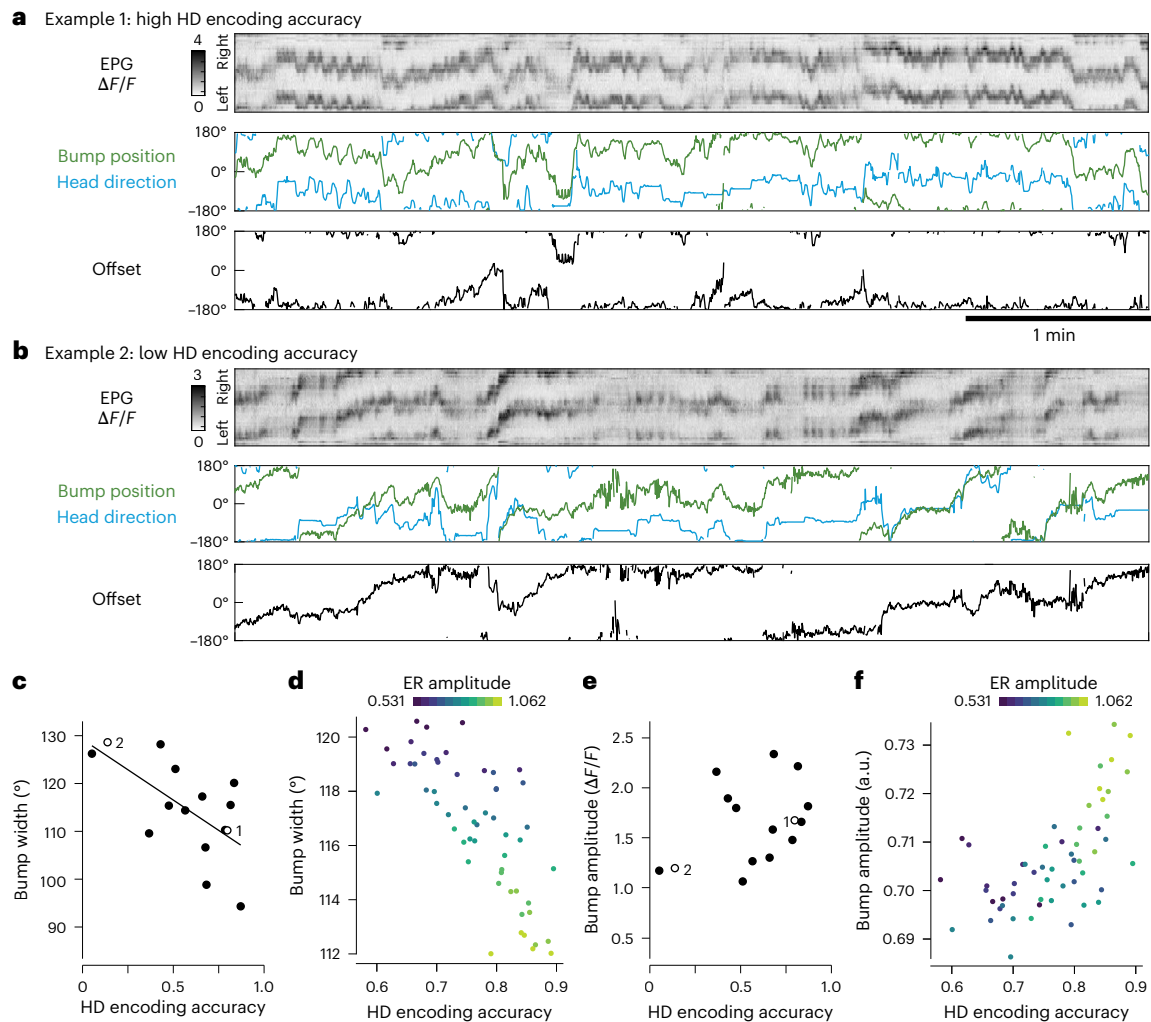
As an aside, we noticed that flies with visually biased HD systems sometimes showed clear behavioral reorientation into the wind after a wind shift (Extended Data Fig. 3). In these flies, the HD system evidently interprets the wind shift as a shift in the wind's environmental direction, not a shift in HD; nonetheless, these flies still reorient into the wind after it shifts. Here, behavioral reorientation into the wind is likely mediated by pathways for orientation control that bypass the HD system<sup>32,33,44</sup>. These observations confirm that these flies are able to detect both cues and the findings are consistent with a model where individual biases originate with ER neurons, as ER neurons are specifically devoted to the HD system.

To summarize, we find that the same cue produces a more accurate HD encoding in some individuals than in others. This implies that the cue carries more or less information for different individuals. Furthermore, each individual accords more weight to the cue they experience as more informative. We can account for these results by positing individual variations in the amplitude of the sensory responses in ER neurons.

### Cue combinations change the bump profile and drive learning

Normally, the relationship between different cues should be relatively stable rather than constantly shifting. As the organism acquires more





**Fig. 3 | The influence of a cue varies across individuals.** This figure shows data for the 15 flies from Fig. 2. **a**, An example fly navigating in the virtual environment with the dim visual cue (HD encoding accuracy = 0.80). The polar histogram shows offset distribution. **b**, Another fly in the same environment (HD encoding accuracy = 0.14). **c**, Individual variations in HD encoding accuracy correlate with bump width ( $P = 0.01$ , Pearson correlation). Data are from the first block where the dim visual cue was presented (200 s in duration). The dim cue produced the

most individual variation (Extended Data Fig. 2). **d**, Model showing that, with the dim cue, small variations in ER amplitude produce strongly correlated changes in bump width and HD encoding accuracy. **e**, Individual variations in HD encoding accuracy do not correlate with bump amplitude ( $P = 0.18$ , Pearson correlation). All data are from the environment with the dim cue. **f**, Model showing that, with a dim cue, small variations in ER amplitude produce weakly correlated changes in bump amplitude and HD encoding accuracy.

familiarity with a given cue, it should ascribe more weight to this cue, as long as the cue continues to appear stable. In principle, this should produce a more accurate HD system and the multisensory environment should be encoded more accurately.

To test these predictions, we combined the two cues in a stable configuration after first testing them individually (Fig. 5a and Extended Data Fig. 4). We found that a stable configuration of these two cues increased HD encoding accuracy (Fig. 5b). This result confirms our expectation that the two cues together convey more information than either cue alone. Moreover, a stable configuration of these two cues also produced a narrower and higher-amplitude bump (Fig. 5b).

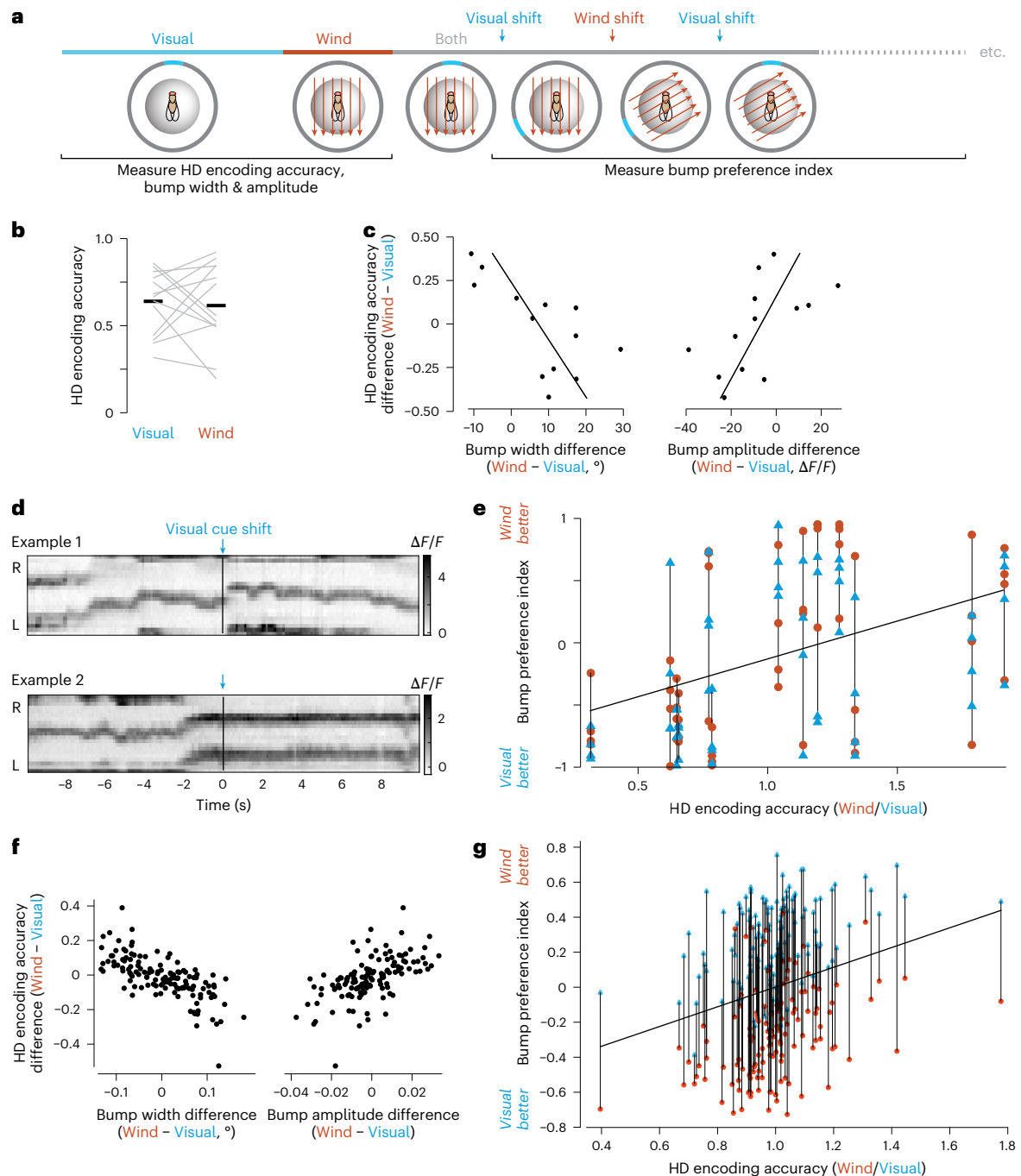
In the two-cue environment, we found that the bump generally retained its offset with respect to the cue that had been presented last (that is, the more recently familiar cue) (Fig. 5c,d). The more recently familiar cue was generally the dominant cue in the two-cue environment, regardless of whether that was the visual cue or the wind (Extended Data Fig. 4). Thus, when we moved the fly into the two-cue environment, its HD system continued to function seamlessly without a change in the representational coding frame.

When we retested each cue at the end of the experiment, we found that the system had reorganized to align the effects of the two cues. In

general, the offset with respect to the less familiar cue (cue 1) changed dramatically, whereas the offset with respect to the familiar cue had changed much less. For example, in the first example shown in Fig. 5c, the wind and the visual cue have initial offsets almost  $180^{\circ}$  apart. Thus, when we place the visual cue in the upwind direction, the two cues are pushing the bump to opposite locations. In this fly, the visual cue was presented last and the bump retained its offset with respect to this cue. At the end of the experiment, we found that the wind offset had changed to match the visual offset.

Overall, we found that the change of each cue offset was predicted by the conflict during the two-cue epoch; here, we define conflict as the difference between the offset with both cues and the initial offset. Specifically, the bump offset often changed dramatically for the less familiar cue where the conflict was often large; conversely, the bump offset generally did not change for the more familiar cue where the conflict was generally small (Fig. 5e). These findings argue that the more familiar cue instructs the remapping of the less familiar cue to produce a self-consistent representation of the environment.

Our network model can explain all these results. When cue 1 appears, the Hebbian learning rule progressively etches a notch into the weights associated with cue 1 (Fig. 5f,g), which slowly increases

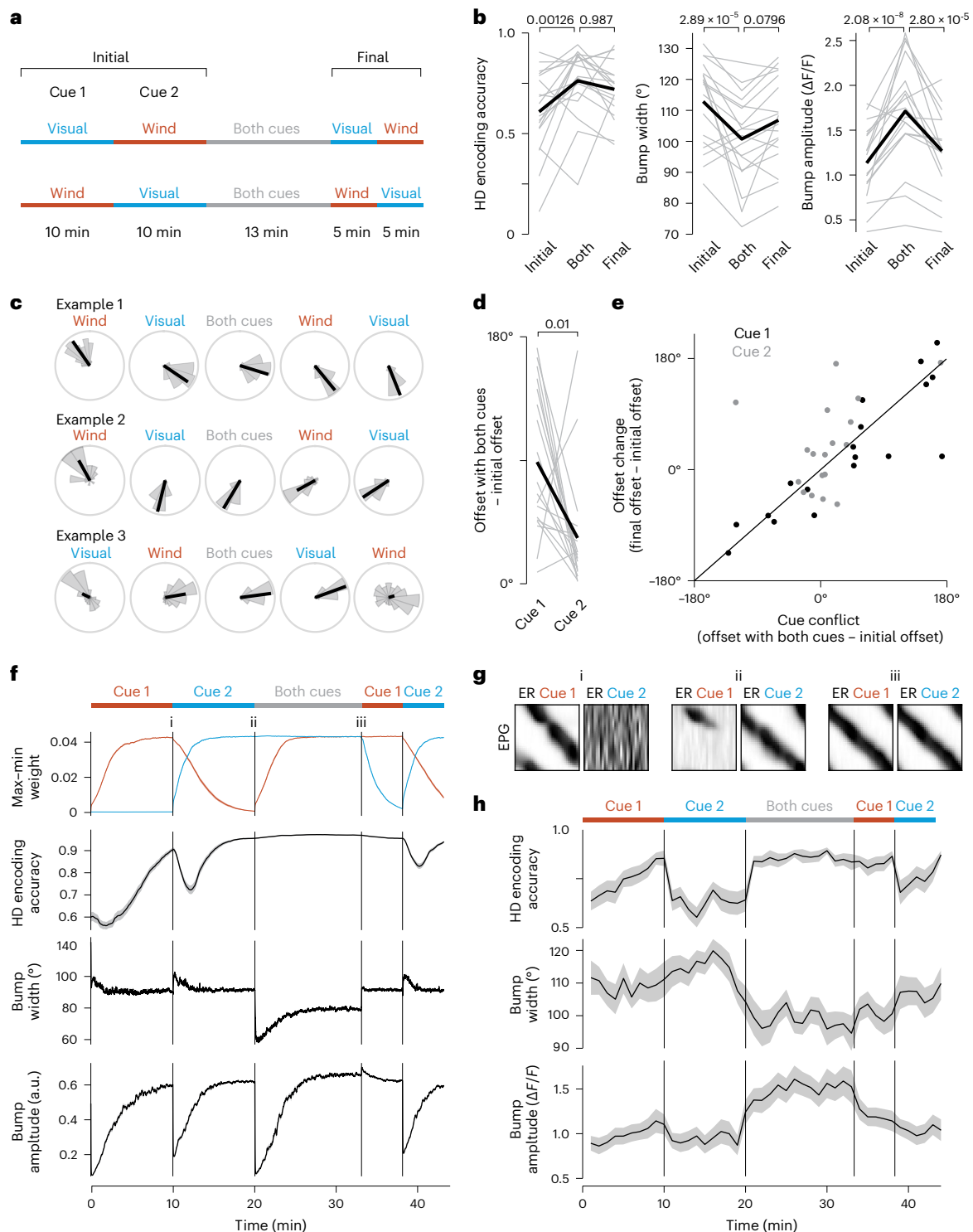


**Fig. 4 | More informative cues are accorded more weight.** This figure shows data for 13 flies. **a**, A bright visual cue and wind cue were tested individually and then combined. One cue was shifted 120° in an alternating sequence every 5 min. In some experiments, the visual cue came first (as shown here) while, in others, the wind came first. **b**, In the initial single-cue epochs, mean HD encoding accuracy is similar for the two cues ( $P = 0.76$ , paired two-sided Wilcoxon test). Thick horizontal lines are mean values. Gray lines are individual flies. **c**, In the initial single-cue epochs, the difference in HD encoding accuracy for the two cues is correlated with the bump width difference ( $R^2 = -0.68$ ,  $P = 0.01$ , Pearson correlation) and the bump amplitude difference ( $R^2 = 0.58$ ,  $P = 0.04$ , Pearson correlation). **d**, Example responses to +120° shifts of the visual cue. In example 1, the bump shifts upward to follow the upward (positive) shift of the cue, whereas,

in example 2, it does not. In both cases, the wind does not shift. Extended Data Figure 3a shows these examples in more detail. **e**, The bump preference index is significantly correlated with relative HD encoding accuracy in the two single-cue environments ( $P = 0.03$ , Pearson correlation). Each gray line at a single encoding accuracy connects the data for one fly (four wind shifts in orange and four visual shifts in blue). An index of -1 means that the bump follows the visual cue, whereas +1 means that the bump follows the wind. This index is close to -1 for example 1 and close to +1 for example 2 (blue numerals). **f, g**, Same as **c** (f) and **e** (g) but for model networks with two populations of ER neurons. In different simulation runs, we varied the amplitude of sensory activity of one ER population while holding the amplitude constant in the other population.

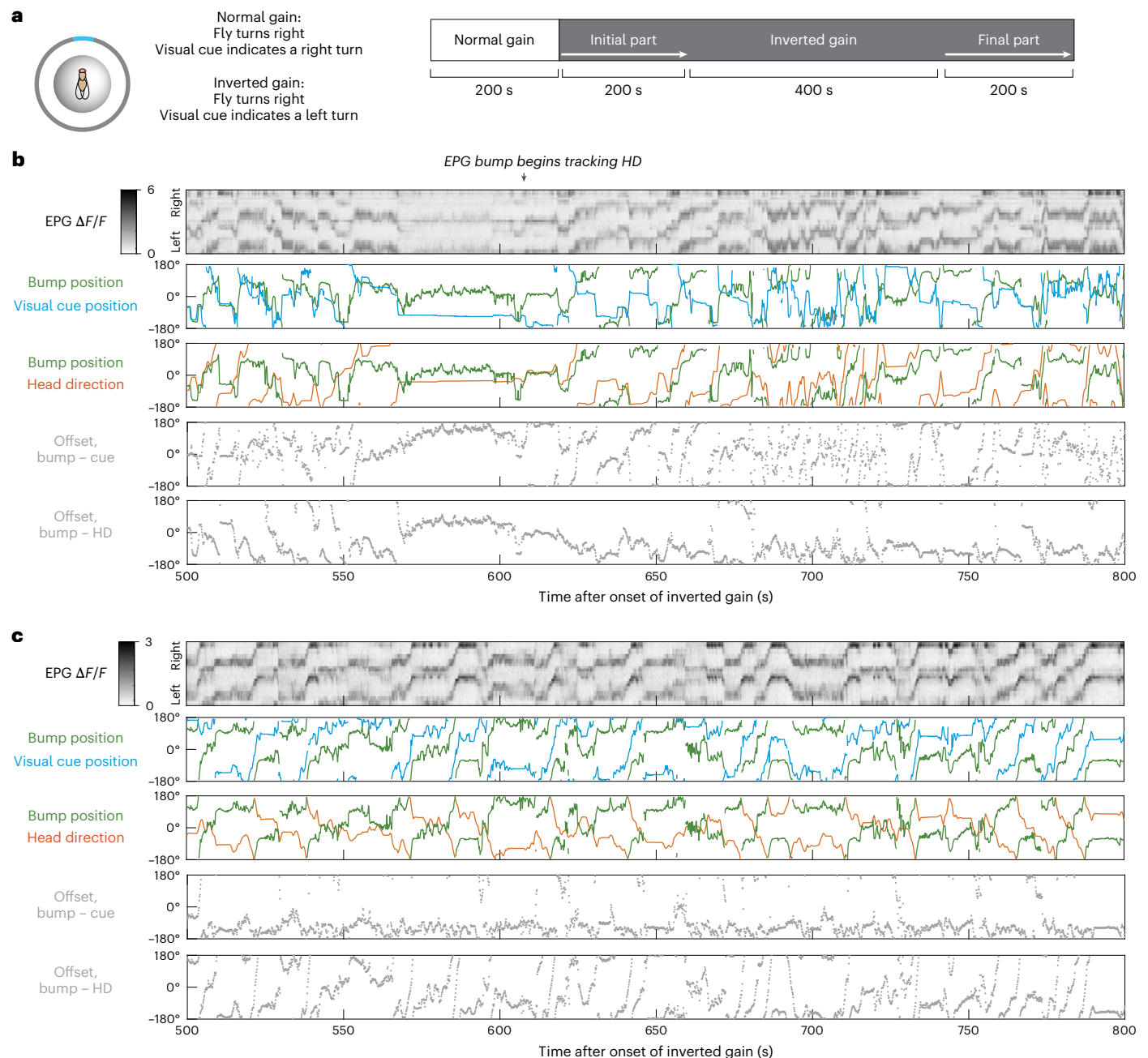
HD encoding accuracy. We see the same slow increase in HD encoding accuracy in our data (Fig. 5h). Note that LTD here represents ‘learning’. Next, cue 2 appears and synaptic plasticity progressively etches a notch in the weights associated with cue 2; meanwhile, the notch

associated with cue 1 is gradually erased, as LTP (‘forgetting’) outweighs LTD (learning) when the cue is absent. This is because LTP depends only on postsynaptic activity. Then, when cue 1 reappears while cue 2 is still present, cue 2 is now dominant, because its weight



**Fig. 5 | Cue combinations change the bump profile and drive learning.** This figure shows data for 18 flies. **a**, Experimental protocol. Each cue was tested alone, then the two cues were combined and, finally, each cue was retested alone. The first cue was either the bright visual cue or wind. During the initial single-cue blocks, the two cues produce similar HD encoding accuracy and similar bump profiles, on average (Extended Data Fig. 4). **b**, HD encoding accuracy, bump width and bump amplitude for all three blocks; *P* values are for linear mixed-effects models with Tukey comparisons and Bonferroni corrections. **c**, Offset distributions for three example flies (Extended Data Fig. 4). **d**, Absolute difference between the offset with both cues and the initial offset. This difference is significantly smaller for cue 2, which is the more familiar cue (*P* value from paired two-sided Wilcoxon test). In most flies, cue 2 captures the bump during

the epoch with both cues. **e**, Offset change (final - initial) versus offset conflict during the cue combination period (both - initial). There are two data points per fly, one for cue 1 (the less familiar cue) and the other for cue 2 (the more familiar cue). The line shown in the plot is a fit to all these points, based on the model that describes the offset change as a linear function of conflict with von Mises noise (Methods). **f**, Model showing the temporal evolution of max-min ER  $\rightarrow$  EPG weight (that is, the range of weight values in the matrix), HD encoding accuracy, bump width and bump amplitude. **g**, Model showing ER  $\rightarrow$  EPG weights at three time points (i, ii and iii). **h**, The same parameters tracked over time in the experimental data (mean  $\pm$  s.e.m. across flies). Here, all parameters were measured in a rolling 60-s time window (Extended Data Fig. 5).



**Fig. 6 | Learning to invert the mapping of a visual cue. a**, Schematic of the protocol. **b**, Data from an example experiment, shown for the last 300 s of the inverted gain block. The EPG bump begins to track the fly's HD (arrowhead).

**c**, Data from another example experiment. In this case, the HD system simply follows the visual cue for the entire inverted gain block and, thus, it does not accurately track HD.

notch is deeper; in other words, the memory of this cue is stronger because it has been familiar more recently (Fig. 5f). In the two-cue environment, the bump is narrow because there is more inhibitory drive to the network. Over time, synaptic plasticity recreates a notch in the weights associated with cue 1 but, now, the Hebbian learning rule ensures that the two notches are well aligned, which explains the offset changes we observed in our data (Fig. 5d,e). This causes the active EPG neurons to be even more disinhibited than they were previously, meaning that bump amplitude increases (Fig. 5g). Subsequently, cue 2 disappears but HD encoding accuracy remains high because cue 1 has been well learned (highly familiar) by this point; this is also a phenomenon we observe in our data (Fig. 5h). This 'priming' effect in the model is because of Hebbian plasticity. This priming effect can account for the observation that a salient cue can persistently increase HD encoding accuracy even after that cue is removed<sup>42</sup>. Lastly, the return to cue 2

produces a drop in HD encoding accuracy because cue 2 has already been partially forgotten.

To summarize, our data show that cue combinations increase HD encoding accuracy, narrow the bump, increase bump amplitude and trigger learning. When two cues conflict, the more familiar cue is weighted more heavily and used to guide the remapping of a less familiar cue. All these findings can be explained by a ring attractor model with highly plastic sensory synapses. In essence, Hebbian plasticity stores the image of a familiar cue and subsequently uses this image to instruct a new round of plasticity at the synapses associated with the less familiar cue. Although we cannot observe synaptic weights in our experimental data, we can infer the pattern of weights by tracking the position and amplitude of the EPG activity bump; specifically, the bump's position (relative to a cue) tells us the location of the notch in the weight matrix, while the bump's amplitude tells us about the notch depth.



### A cue that produces a wide bump is remapped more readily

There is a basic tradeoff between stability and flexibility in any representation of the environment; when the representation is stable, it is accurate but this makes remapping more difficult. In particular, if the HD system ascribes too much weight to any particular cue, it may be difficult to learn a new interpretation of that cue. We can state this idea in mechanistic terms; if some sensory connections onto HD cells become dominant, then other inputs cannot compete with them and their weights cannot change. Therefore, a cue that is weighted heavily may be difficult to remap after the environment changes.

To test this prediction, we challenged flies to learn an inverted gain in virtual reality, such that the visual cue moves in the ‘wrong’ direction whenever the fly rotates (Fig. 6a). This is a dramatic change in the environment, which produces a strong conflict between the visual cue and the self-motion input to the HD system. In principle, it might seem that the HD system should simply ignore the visual cue in this situation but this would lead to poor HD encoding accuracy because this system requires feedback for accurate performance (Fig. 2c). Therefore, the network should ideally learn to invert the mapping of the visual cue onto the EPG cell ensemble. A previous study reported that an optogenetic method could be used to artificially create this type of inversion<sup>22</sup> and this motivated us to investigate whether it was possible to obtain the same type of inversion through visuomotor learning.

Notably, we found that some individuals were able to learn to invert the mapping of the visual cue. In these individuals, bump dynamics were generally unpredictable during the initial part of the inverted gain experience; however, at some point, the bump would begin tracking HD fairly accurately (Fig. 6b). By contrast, other individuals never learned to invert their interpretation of the visual cue. In these cases, the bump generally just tracked the visual cue; thus, when the fly turned right, the HD system registered a left turn (Fig. 6c).

To quantify learning, we defined a remapping index, where +1 means that the bump is correctly tracking the fly’s rotation, implying successful remapping; conversely, −1 means that the bump is moving against the fly’s rotation, implying no remapping (Fig. 7a). On average, the remapping index was significantly higher in the final part of the inverted gain block as compared to the initial part (Fig. 7b). Moreover, during the inverted gain block, the mean HD encoding accuracy increased (Fig. 7c) and the consistency of behavioral orienting also increased (Fig. 7d). Thus, spatial learning was clearly occurring during the inverted gain block.

Importantly, the individuals with high remapping indices were not simply learning to ignore the visual cue because their HD representation was much more stable than it would be in darkness. In darkness, HD encoding accuracy was always low (Fig. 2c), whereas HD encoding accuracy was relatively high by the final part of the inverted gain block, at least in some individuals (Fig. 7c). Thus, these individuals are not learning to discount the cue; rather, they are clearly learning to invert their interpretation of the cue.

We could predict which individuals would be successful at learning this inversion on the basis of the profile of the EPG bump in the normal gain block before the onset of inverted gain. Specifically, we found that individuals with the widest bumps (in the normal gain block) all had relatively high remapping indices at the end of the inverted gain block (Fig. 7e). In other words, a wider bump was predictive of better remapping. There was also a trend for these same individuals to have lower HD encoding accuracy during the normal gain block, consistent with their wider bumps, although this fell just short of statistical significance (Extended Data Fig. 6). Bump amplitude in the normal gain block was not correlated with remapping (Fig. 7f).

Similarly, we found that our network model could also learn to invert its interpretation of the visual cue. This works because self-motion cues instruct Hebbian learning at ER → EPG synapses; thus, ER neurons can learn to push the bump in whatever direction is consistent with self-motion signals. In the model, we can obtain

individual variations in performance by varying the amplitude of visually evoked activity in ER neurons (Figs. 3 and 4). We found that a lower ER amplitude allows faster remapping (Fig. 7g) because self-motion cues have more weight relative to the visual cue. This allows self-motion cues to better instruct the appropriate reinterpretation of the visual cue, thus more rapidly reversing the orientation of the notch in the ER → EPG weight matrix (Fig. 7h). A lower ER amplitude also produces a wider bump (Fig. 7g), which explains why a wider bump predicts faster learning (Fig. 7e). In the model, a lower ER amplitude also produces a smaller bump amplitude, which was not correlated with learning in our data; this may reflect our limited sample size or a real difference between the model and the actual network.

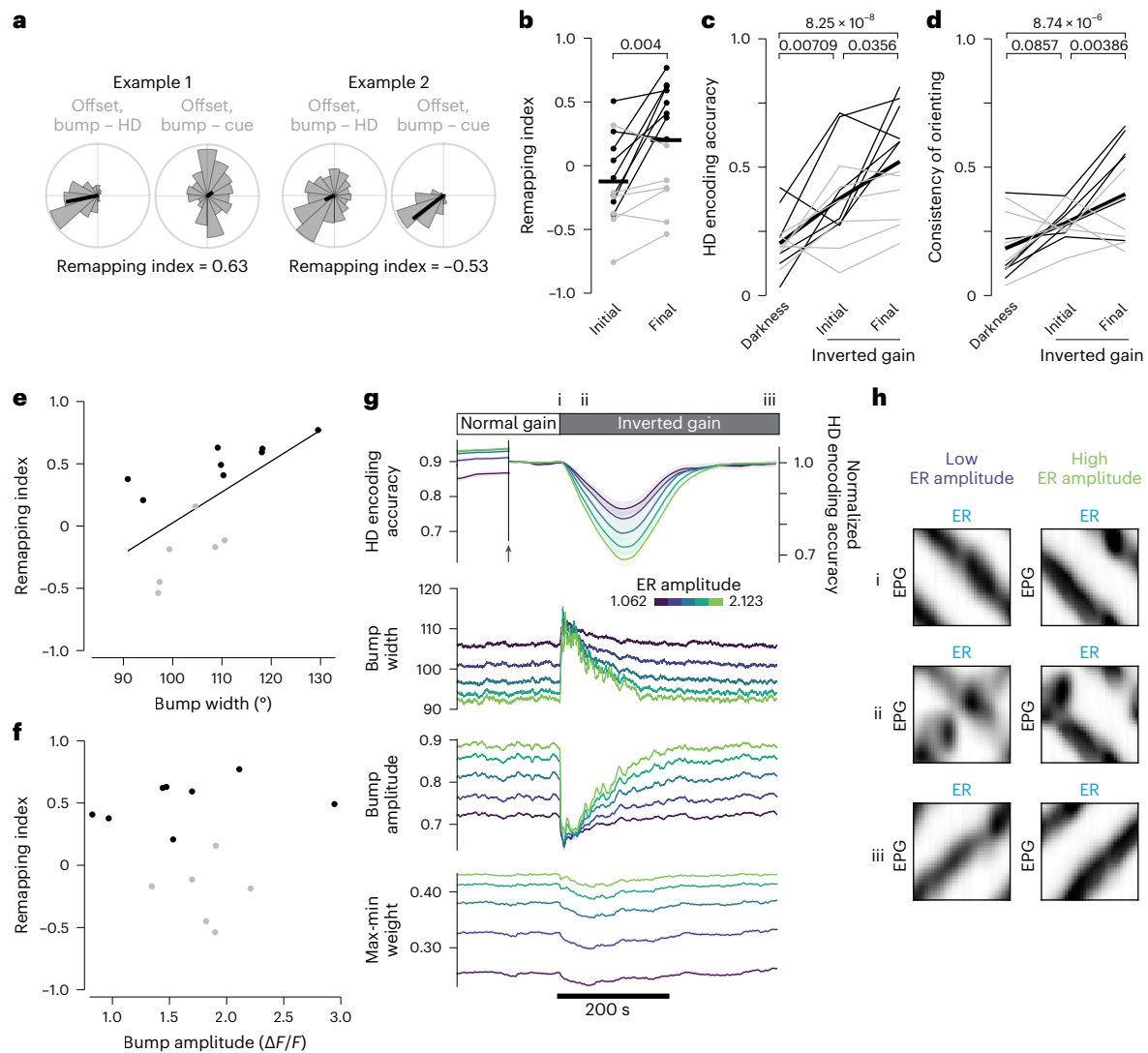
In summary, we found that, when a cue produces a wider bump, it is remapped more readily after a switch to inverted gain. We can reproduce this result in a network model where we vary the amplitude of the sensory inputs to the HD network. In this model, individuals that experience the cue as being less intense are better able to remap that cue after a switch to inverted gain. This result highlights a basic tradeoff between stability and flexibility in this system; when a spatial representation is highly stable, it is highly accurate but this also makes it more difficult to adjust this representation when conditions change.

## Discussion

There is a tradeoff between stability and flexibility in neural network dynamics. On one hand, theoretical attractor networks are often tuned to be stable, such that the bump of activity in the network hardly varies in amplitude or width<sup>28,45,46</sup>; these networks are designed to support working memory but not learning. Theoretical attractor networks are sometimes also assumed to have fixed synaptic weights at their sensory inputs to minimize representational drift<sup>47,48</sup>. On the other hand, the biological attractor networks that underpin the brain’s navigational brain regions<sup>17</sup> likely support learning about the relative positions of environmental cues and the stability of those cues. Thus, as cues change and learning proceeds, we might expect to see changes in synaptic weights<sup>20–28</sup> and resulting changes in profile<sup>49,50</sup>. We do not have a general understanding of how biological neural networks manage the tradeoff between stability and flexibility.

In the *Drosophila* HD system, the bump of activity has a stable position when the fly is not moving, effectively storing a working memory of the fly’s orientation, using a ring attractor network. Here, we show that this bump varies systematically in width and amplitude depending on the intensity and familiarity of the cues in the environment. The changes we observe in the bump profile can be explained, in part, by changes in synaptic weights at sensory synapses onto HD cells. Although we cannot observe these synaptic weights directly, we can ascertain something about these weights on the basis of observed changes in bump profile. For example, when cue intensity increases or a cue becomes more familiar, we find that the bump becomes narrower and this can be at least partly explained by a deeper notch in the synaptic weight matrix. Moreover, we show that increases in cue intensity and familiarity produce increases in HD encoding accuracy and the consistency of behavioral orientation; thus, a more intense and familiar cue is actually more informative. When we then introduce conflicts between cues, we find that the cue that produces higher encoding accuracy is generally accorded more weight and this is predicted by the bump profile in response to that cue. Moreover, the cue that produces higher encoding accuracy generally instructs spatial learning in response to the cue conflict, which resolves the conflict and increases encoding accuracy. Together, these results provide concrete evidence for the theoretical proposal<sup>20,21,26–31</sup> that Hebbian plasticity can endow an attractor with the flexibility it needs to support learning about the relative positions of environmental cues and the stability of those cues.

The idea that spatial learning might involve synaptic plasticity is intuitive but it is technically difficult to observe synaptic weights evolve in real time in a biological network. Here, we took the bump



**Fig. 7 | Bump width in normal gain predicts cue remapping in inverted gain.**

Panels **b**, **e**, **f** show data from 14 flies, while **c**, **d** show data from 12 flies because two flies did not receive a 'darkness' trial. **a**, Offset distributions for the two example flies from Fig. 6. The remapping index is positive for fly 1 because the offset with respect to HD is more stable. The remapping index is negative for fly 2 because the offset with respect to the visual cue is more stable. **b**, Remapping index in the initial and final part of the inverted gain block. Across all flies, the remapping index is larger in the final part ( $P$  value from two-sided Wilcoxon test). Black and gray lines denote individuals with final remapping indices above or below the mean (horizontal bar); this is preserved in **c**–**f**. **c**, **d**, HD encoding accuracy (**c**) and consistency of behavioral orientation (**d**) in darkness versus inverted gain ( $P$  values from linear mixed-effects models with Tukey comparisons and Bonferroni corrections). **e**, **f**, Final remapping index is predicted by bump width (**e**;  $P = 0.02$ , Pearson's correlation) but not bump amplitude (**f**;  $P = 0.79$ , Pearson's

correlation) during the preceding normal gain block ( $n = 14$  flies) (Extended Data Fig. 6). **g**, Model showing the effect of varying the amplitude of visual responses in ER neurons. Lower ER amplitude produces a faster recovery of HD encoding accuracy after the onset of inverted gain. Note that lower ER amplitude produces lower overall HD encoding accuracy; to facilitate a comparison between conditions, we show the normalized HD encoding accuracy (normalized to 1 at end of normal gain epoch) beginning halfway through the normal gain epoch (arrow). Results are shown as the mean  $\pm$  s.e.m. across 100 simulation runs. Panel **h** shows weight matrices for the indicated time points (i, ii and iii). **h**, Model showing ER  $\rightarrow$  EPG weights at the end of the normal gain block (i), soon after the onset of inverted gain (ii) and when remapping is nearly complete (iii). Results are shown for two ER amplitudes (1.327 and 2.123). At the intermediate time point (ii), remapping has progressed further with the lower ER amplitude.

profile as a proxy for synaptic weights and showed that it takes minutes to achieve restabilization of the bump profile after a change in the environment; this result implies that it also takes minutes to achieve restabilization of synaptic weights. This timescale may represent a compromise between the demands of flexibility and stability; a faster timescale would accelerate learning but it would also accelerate forgetting when a cue transiently disappears and increase representational drift. In the *Drosophila* HD system, dopamine controls the learning rate and dopamine increases with exploratory movements<sup>24</sup>; this provides a mechanism to increase network flexibility during exploration, conversely allowing more stability at other times.

It is less intuitive to think that synaptic plasticity can help cues to be weighted by their informativeness. Accurate inference requires that more informative cues should be weighted more heavily and, in some cases, this reweighting process can occur instantaneously; for example, we immediately ascribe less weight to a visual cue when it is blurred or degraded<sup>51,52</sup>. In these cases, there is no need for inference to involve neural plasticity. However, in other cases, the informativeness of a cue cannot be assessed immediately; for example, during navigation, the informativeness of a visual cue depends on its stability within the environment, which can only be evaluated by learning. Hebbian plasticity has been suggested to assist inference under these conditions<sup>21,25,29–31</sup> and our work provides direct support for this suggestion.

Importantly, our model can explain many features of animal navigation behavior. For example, navigating animals assign more weight to spatial cues that are more salient or more stable and familiar; this is true in both insects<sup>8–13</sup> and vertebrates<sup>3–7,14</sup>. Our results imply that these types of cues can create stronger spatial modulations in synaptic weights from sensory cells. Previous work has also demonstrated that insects can form arbitrary learned associations between the angular positions of the wind and the sun<sup>8,9,27</sup>. Our results show how an arbitrary cue configuration can be stored in the pattern of synaptic weights from sensory cells.

Our results also provide insight into mechanisms underlying individual differences in navigation behavior<sup>53,54</sup>. Specifically, we showed that different individuals have different levels of encoding accuracy in the same virtual reality environment. Moreover, individuals with higher accuracy had significantly narrower bumps of activity. On the other hand, individuals with wider bumps were better able to reorganize their HD representations when the environment changes. In our model, we could recapitulate all these individual differences by varying the amplitude of sensory inputs to the HD system. Importantly, individual variation is not specific to flies; human subjects also show marked idiosyncratic differences in cue weighting during navigation<sup>14,55–58</sup>. Individual differences may reflect variations in each individual's past experiences. Alternatively, they may reflect an evolutionary strategy; for example, it might be useful for the species if some individuals have more stable mental representations of space (sacrificing flexibility), while others have more flexible representations (sacrificing stability).

The fact of individual differences in navigation should make it obvious that navigation is not always optimized. Although the model network we describe here has some remarkable features, it cannot perform optimal Bayesian inference, even in the absence of these individual differences. Optimal inference would require the network to store the certainty associated with the network's HD estimate, for example through some mechanism of persistent activity that boosts bump amplitude after an informative cue is presented, such that bump amplitude remains persistently high for some time after the cue disappears<sup>49,50</sup>. Our data imply that this does not occur (Extended Data Fig. 8); accordingly, our model network in this study does not store certainty about HD, instead storing information about the salience, stability and familiarity of each cue. Curiously, there are hints that the insect brain does have a way to represent certainty during navigation. For example, homing ants will search for their nest over a wider radius if they have just returned from a longer trip, which suggests that they keep track of their certainty and use this to adjust their search strategy<sup>59</sup>. Moreover, ants will steer further downwind of their expected nest site when they have returned from a longer trip, which should help them to use odor filaments as a guideline back to their nest when their certainty is low<sup>60</sup>. In the future, it will be interesting to investigate whether these behaviors actually arise from a neural representation of certainty or whether they reflect simpler behavioral strategies<sup>27</sup>.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-024-01823-z>.

## References

- Cheng, K., Shettleworth, S. J., Huttenlocher, J. & Rieser, J. J. Bayesian integration of spatial information. *Psychol. Bull.* **133**, 625–637 (2007).
- Durrant-Whyte, H. & Bailey, T. Simultaneous localization and mapping: part I. *IEEE Robot. Autom. Mag.* **13**, 99–110 (2006).
- Etienne, A. S., Teroni, E., Hurni, C. & Portenier, V. The effect of a single light cue on homing behaviour of the golden hamster. *Anim. Behav.* **39**, 17–41 (1990).
- Biegler, R. & Morris, R. G. Landmark stability: further studies pointing to a role in spatial learning. *Q. J. Exp. Psychol. B* **49**, 307–345 (1996).
- Biegler, R. & Morris, R. G. Landmark stability is a prerequisite for spatial but not discrimination learning. *Nature* **361**, 631–633 (1993).
- Sutton, J. E. Multiple-landmark piloting in pigeons (*Columba livia*): landmark configuration as a discriminative cue. *J. Comp. Psychol.* **116**, 391–403 (2002).
- Shettleworth, S. J. & Sutton, J. E. Multiple systems for spatial learning: dead reckoning and beacon homing in rats. *J. Exp. Psychol. Anim. Behav. Process.* **31**, 125–141 (2005).
- Muller, M. & Wehner, R. Wind and sky as compass cues in desert ant navigation. *Naturwissenschaften* **94**, 589–594 (2007).
- Dacke, M. et al. Multimodal cue integration in the dung beetle compass. *Proc. Natl Acad. Sci. USA* **116**, 14248–14253 (2019).
- Wystrach, A., Mangan, M. & Webb, B. Optimal cue integration in ants. *Proc. Biol. Sci.* **282**, 20151484 (2015).
- Legge, E. L. G., Wystrach, A., Spetch, M. L. & Cheng, K. Combining sky and earth: desert ants (*Melophorus bagoti*) show weighted integration of celestial and terrestrial cues. *J. Exp. Biol.* **217**, 4159–4166 (2014).
- Narendra, A., Gourmaud, S. & Zeil, J. Mapping the navigational knowledge of individually foraging ants, *Myrmecia croslandi*. *Proc. Biol. Sci.* **280**, 20130683 (2013).
- Shaverdian, S. et al. Weighted cue integration for straight-line orientation. *iScience* **25**, 105207 (2022).
- Chen, X., McNamara, T. P., Kelly, J. W. & Wolbers, T. Cue combination in human spatial navigation. *Cogn. Psychol.* **95**, 105–144 (2017).
- Taube, J. S. & Burton, H. L. Head direction cell activity monitored in a novel environment and during a cue conflict situation. *J. Neurophysiol.* **74**, 1953–1971 (1995).
- Jeffery, K. J. Learning of landmark stability and instability by hippocampal place cells. *Neuropharmacology* **37**, 677–687 (1998).
- Knierim, J. J., Kudrimoti, H. S. & McNaughton, B. L. Place cells, head direction cells, and the learning of landmark stability. *J. Neurosci.* **15**, 1648–1659 (1995).
- Knierim, J. J. & Zhang, K. Attractor dynamics of spatially correlated neural activity in the limbic system. *Annu. Rev. Neurosci.* **35**, 267–285 (2012).
- Touretzky, D. S. in *Head Direction Cells and the Neural Mechanisms of Spatial Orientation* (eds Wiener, S. I. & Taube, J. S.) Ch. 18 (MIT Press, 2005).
- Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S. & McNaughton, B. L. A model of the neural basis of the rat's sense of direction. *Adv. Neural Inf. Process. Syst.* **7**, 173–180 (1995).
- Page, H. J. I. et al. A theoretical account of cue averaging in the rodent head direction system. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130283 (2014).
- Kim, S. S., Hermundstad, A. M., Romani, S., Abbott, L. F. & Jayaraman, V. Generation of stable heading representations in diverse visual scenes. *Nature* **576**, 126–131 (2019).
- Fisher, Y. E., Lu, J., D'Alessandro, I. & Wilson, R. I. Sensorimotor experience remaps visual input to a heading-direction network. *Nature* **576**, 121–125 (2019).
- Fisher, Y., Marquis, M., D'Alessandro, I. & Wilson, R. I. Dopamine promotes head direction plasticity during orienting movements. *Nature* **612**, 316–322 (2022).
- Knight, R. et al. Weighted cue integration in the rodent head direction system. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20120512 (2014).



26. Cope, A. J., Sabo, C., Vasilaki, E., Barron, A. B. & Marshall, J. A. A computational model of the integration of landmarks and motion in the insect central complex. *PLoS ONE* **12**, e0172325 (2017).
27. Mitchell, R., Shaverdian, S., Dacke, M. & Webb, B. A model of cue integration as vector summation in the insect brain. *Proc. Biol. Sci.* **290**, 20230767 (2023).
28. Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* **16**, 2112–2126 (1996).
29. Mulas, M., Waniek, N. & Conradt, J. Hebbian plasticity realigns grid cell activity with external sensory cues in continuous attractor models. *Front. Comput. Neurosci.* **10**, 13 (2016).
30. Keinath, A. T., Epstein, R. A. & Balasubramanian, V. Environmental deformations dynamically shift the grid cell spatial metric. *eLife* **7**, e38169 (2018).
31. Ocko, S. A., Hardcastle, K., Giocomo, L. M. & Ganguli, S. Emergent elasticity in the neural code for space. *Proc. Natl Acad. Sci. USA* **115**, E11798–E11806 (2018).
32. Giraldo, Y. M. et al. Sun navigation requires compass neurons in *Drosophila*. *Curr. Biol.* **28**, 2845–2852.e4 (2018).
33. Green, J., Vijayan, V., Mussells Pires, P., Adachi, A. & Maimon, G. A neural heading estimate is compared with an internal goal to guide oriented navigation. *Nat. Neurosci.* **22**, 1460–1468 (2019).
34. Seelig, J. D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521**, 186–191 (2015).
35. Hulse, B. K. et al. A connectome of the *Drosophila* central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife* **10**, e66039 (2021).
36. Kim, S. S., Rouault, H., Druckmann, S. & Jayaraman, V. Ring attractor dynamics in the *Drosophila* central brain. *Science* **356**, 849–853 (2017).
37. Turner-Evans, D. et al. Angular velocity integration in a fly heading circuit. *eLife* **6**, e23496 (2017).
38. Green, J. et al. A neural circuit architecture for angular integration in *Drosophila*. *Nature* **546**, 101–106 (2017).
39. Okubo, T. S., Patella, P., D'Alessandro, I. & Wilson, R. I. A neural network for wind-guided compass navigation. *Neuron* **107**, 924–940 (2020).
40. Dana, H. et al. High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat. Methods* **16**, 649–657 (2019).
41. Wolff, T., Iyer, N. A. & Rubin, G. M. Neuroarchitecture and neuroanatomy of the *Drosophila* central complex: a GAL4-based dissection of protocerebral bridge neurons and circuits. *J. Comp. Neurol.* **523**, 997–1037 (2015).
42. Haberkern, H. et al. Maintaining a stable head direction representation in naturalistic visual environments. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.05.17.492284> (2022).
43. Seelig, J. D. & Jayaraman, V. Feature detection and orientation tuning in the *Drosophila* central complex. *Nature* **503**, 262–266 (2013).
44. Rayshubskiy, A. et al. Neural circuit mechanisms for steering control in walking *Drosophila*. *eLife* **13**, RP102230 (2025).
45. Wang, X. J. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* **24**, 455–463 (2001).
46. Compte, A., Brunel, N., Goldman-Rakic, P. S. & Wang, X. J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* **10**, 910–923 (2000).
47. Xie, X., Hahnloser, R. H. & Seung, H. S. Double-ring network model of the head-direction system. *Phys. Rev. E* **66**, 041902 (2002).
48. Redish, A., Elga, A. & Touretzky, D. A coupled attractor model of the rodent head direction system. *Network* **7**, 671–685 (1996).
49. Kutschireiter, A., Basnak, M. A. & Drugowitsch, J. Bayesian inference in ring attractor networks. *Proc. Natl Acad. Sci. USA* **120**, e2210622120 (2022).
50. Carroll, S., Josić, K. & Kilpatrick, Z. P. Encoding certainty in bump attractors. *J. Comput. Neurosci.* **37**, 29–48 (2014).
51. Alais, D. & Burr, D. The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* **14**, 257–262 (2004).
52. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
53. Ayroles, J. F. et al. Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proc. Natl Acad. Sci. USA* **112**, 6706–6711 (2015).
54. Buchanan, S. M., Kain, J. S. & de Bivort, B. L. Neuronal control of locomotor handedness in *Drosophila*. *Proc. Natl Acad. Sci. USA* **112**, 6700–6705 (2015).
55. Scherer, J. et al. Not seeing the forest for the trees: combination of path integration and landmark cues in human virtual navigation. *Front. Behav. Neurosci.* **18**, 1399716 (2024).
56. Weisberg, S. M. & Newcombe, N. S. How do (some) people make a cognitive map? Routes, places, and working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **42**, 768–785 (2016).
57. Blajenkova, O., Motes, M. A. & Kozhevnikov, M. Individual differences in the representations of novel environments. *J. Environ. Psychol.* **25**, 97–109 (2005).
58. Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F. & Epstein, R. A. Variations in cognitive maps: understanding individual differences in navigation. *J. Exp. Psychol. Learn. Mem. Cogn.* **40**, 669–682 (2014).
59. Merkle, T., Knaden, M. & Wehner, R. Uncertainty about nest position influences systematic search strategies in desert ants. *J. Exp. Biol.* **209**, 3545–3549 (2006).
60. Wolf, H. & Wehner, R. Desert ants compensate for navigation uncertainty. *J. Exp. Biol.* **208**, 4223–4230 (2005).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025



## Methods

### Experimental model and subject details

*Drosophila melanogaster* were raised on cornmeal and molasses (Archon Scientific) under a 12-h light–dark cycle at 25 °C. Experiments were performed on 1-day-old virgin females with the genotype *w/+;+;P{R60D05-Gal4}attP2/P{20XUAS-IVS-jGCaMP7f}VK00005*. Both *P{R60D05-Gal4}attP2* and *P{20XUAS-IVS-jGCaMP7f}VK00005* were obtained from the Bloomington *Drosophila* Stock Center (RRID:BDSC\_39247 and RRID:BDSC\_79031, respectively). *P{R60D05-Gal4}attP2* drives Gal4 expression in EPG neurons, as reported previously<sup>34,41</sup>, and its construction was described previously<sup>61</sup>. *P{20XUAS-IVS-jGCaMP7f}VK00005* was also described previously<sup>40</sup>.

### Fly selection and housing

Virgin female flies were anesthetized on CO<sub>2</sub>, collected at least 12 h before the experiment and then allowed to recover on molasses food. Imaging experiments were conducted the following day, 14–36 h after eclosion. Before the experiment, flies were starved for 0–24 h on a piece of damp laboratory tissue (Kimtech). The starvation time was chosen on the basis of our observations for what resulted in the best fly behavior during the months in which each experiment was conducted (Figs. 2 and 3 and Extended Data Figs. 2, 5, 7 and 8: 5–24 h; Fig. 4 and Extended Data Fig. 3: 0 h; Fig. 5 and Extended Data Fig. 4: 0–3 h; Figs. 6 and 7 and Extended Data Fig. 6: 18–27 h). No statistical methods were used to predetermine sample sizes but our sample sizes are similar to those reported in previous publications<sup>23,24,33,34,37,38</sup>.

### Fly preparation and dissection

Flies were briefly cold-anesthetized in a glass vial (V7005-500EA, Sigma-Aldrich) on ice and placed inside a custom-made inverted pyramidal platform CNC machined from black Delrin (Protolabs Inc.). The head was tilted forward to make the posterior part of the brain more accessible during imaging. Because of the head's angle and the pyramidal shape of the holder, the majority of each eye was positioned below the holder and, therefore, able to see the visual stimuli. The wings were removed and the head and thorax were secured to the holder using ultraviolet-curable glue (Loctite AA 3972) and cured with ultraviolet light (LED-200, Electro-Lite). To prevent large brain movements, the proboscis was removed (we again briefly cold-anesthetized the animal during this process). The head was then bathed in an extracellular saline solution with the following composition: 103 mM NaCl, 3 mM KCl, 5 mM TES, 8 mM trehalose, 10 mM glucose, 26 mM NaHCO<sub>3</sub>, 1 mM NaH<sub>2</sub>PO<sub>4</sub>, 1.5 mM CaCl<sub>2</sub> and 4 mM MgCl<sub>2</sub> (osmolarity 270–275 mOsm, bubbled with 95% O<sub>2</sub> and 5% CO<sub>2</sub> to reach a final pH of ~7.3). A window was opened in the head cuticle and the trachea and fat were removed to better expose the brain. To reduce additional brain movements, muscle 16 was clipped.

### Two-photon calcium imaging

We performed in vivo calcium imaging with a two-photon laser scanning microscope with a galvo-resonant scanner (Bergamo II, Thorlabs). We used a femtosecond Ti:sapphire laser with dispersion precompensation (Vision-S, Coherent) tuned to 940 nm to achieve two-photon excitation. To image, we used a ×20 objective (numerical aperture: 1.0; XLUMPLN, Olympus) mounted on a fast objective scanner (P-725, Physik Instrumente). The emission from our samples was detected with a GaAsP photomultiplier tube (PMT) detector (Hamamatsu) equipped with a 525-nm bandpass filter (Thorlabs). We collected imaging data using National Instruments PXIe-6341 hardware with ScanImage<sup>62</sup> 2018b or 2020 (Vidrio Technologies, RRID:SCR\_014307). We defined an image as 256 × 128 pixels encompassing the protocerebral bridge (PB). We acquired a volume of 12 slices, with 5-μm steps separating consecutive slices, at a rate of 9.18 volumes per second. We discarded the slices corresponding to 'flyback' frames post hoc.

### Fly locomotion

The platform holding the fly was positioned above a spherical treadmill, consisting of a 9-mm-diameter ball made of foam (FR-4615, General Plastics). The ball was floated on a steady stream of medical grade breathing air at ~0.2–0.3 L min<sup>-1</sup> in a custom holder three-dimensionally (3D) printed using Grey Pro v2 resin (Formlabs). An irregular black pattern was painted on the ball (Vallejo Black Model Color Paint) to allow tracking of the ball surface with machine vision. The fly was positioned on the ball under visual control, using a side camera (CM3-U3-13Y3M-CS, forward-looking infrared (FLIR); InfiniStix lens 94 mm, ×0.50, Infinity Photo-Optical) and a front camera (BFS-U3-13Y3M, FLIR; InfiniStix lens 94 mm, ×0.50, Infinity Photo-Optical). The ball was illuminated with a round board of 36 infrared light-emitting diodes (LEDs) (SODIAL). The image from one of the cameras was acquired at 50 Hz and analyzed using FicTrac<sup>63</sup> version 2.1.1 ([rjdmoore.net/fictrac/](http://rjdmoore.net/fictrac/)) to track the position of the ball in the pitch, yaw and roll axes, thereby reconstructing the fly's locomotor trajectory. This camera was positioned at an angle to ensure that a sufficient fraction of the ball surface was always visible, avoiding occlusion by the rotating air nozzle used to deliver the wind.

### Visual stimuli

Visual stimuli were displayed on a custom-built cylindrical panorama of LEDs adapted from published prototypes<sup>64</sup>. The panorama covered the entire 360° range of azimuthal angles. It consisted of two rows and 12 columns of square blue LED panels (peak: 470 nm), with each panel consisting of 8 × 8 pixels. One panel was removed in the top row (on the fly's right) to accommodate the side camera used for fly positioning and ball tracking. The visual arena was tilted forward to match the inclination of the fly's head in the platform.

The LED panels were covered with a diffuser material (SXF-0600, snow-white light diffuser, Decorative Films). Moreover, five layers of gel filters were used to reduce overlap in spectra and to decrease the intensity of the stimuli: three layers of Tokyo blue (Rosco, RE071), one layer of 0.3 neutral density (Rosco, RE209) and one layer of marine blue (Rosco, RE131). Additionally, the back, top and bottom of the panorama were covered with black tape to further reduce the amount of LED light reaching the PMTs.

The visual cue consisted of a blue vertical stripe (two pixels wide, 7.5°), spanning the vertical extent of the panorama. Visual stimuli were programmed in Matlab 2020a (MathWorks, RRID:SCR\_001622). Custom Python software was used to read FicTrac outputs and generate analog voltage signals through a Phidget analog output device (Phidget Analog 4-Output 1002\_0B). For closed-loop control of the stimulus, the ball displacement in the yaw axis was used to update the azimuthal position of the visual cue (refresh rate ≥ 372 Hz). Analog output signals from the visual panel system were digitized with a NI-DAQ PCI-6351 (National Instruments) at 4 kHz. The intensity of the background was 0/15. The bright cue had an intensity of 15/15, whereas the dim cue had an intensity of 1/15. These intensity values were chosen empirically on the basis of the results of pilot experiments, to ensure that HD encoding accuracy was lower for the dim cue versus the bright cue, but HD encoding accuracy was still higher for the dim cue than in conditions of darkness. We also deliberately chose values of cue brightness and wind speed (discussed below) such that the bright cue and the wind were roughly equally informative, meaning that they produced similar values of HD encoding accuracy (averaged across flies).

### Wind stimuli

Wind stimuli were delivered using a custom-built device that is conceptually similar to a previously published device<sup>65</sup>. Our device uses a commutator to maintain the air flow as the nozzle rotates around the fly in 360°. The commutator was 3D-printed from Rigid 4000 resin (Formlabs) in two pieces. The base piece had a ball holder and two air intake ports (one for the air-supported ball and another for the wind

delivery). The top piece had an air nozzle (inner diameter: 2.8 mm) that rotated around the fly. When the nozzle was in front of the fly, it was positioned 10 mm from the antennae. A ball bearing (McMaster-Carr, 5908K19) was used to create a smoothly rotating interface between the top and bottom pieces. To rotate the air nozzle, a timing belt (0.25-inch width; McMaster-Carr, 6484K118) was attached to the top part of the commutator and connected to a pinion pulley (Servocity SKU, 615424), which was mounted on the shaft of a stepper motor (Pololu, 1204). The entire wind delivery device was designed such that it would fit inside the 360° visual panorama described above.

The air nozzle was printed from black material (Black Resin, Formlabs) to reduce its visibility. Control experiments were performed to ensure that the nozzle did not interfere with the fly's ability to see the LED arena and that the nozzle itself was not acting as a visual cue. In these control experiments, each fly received three stimulus blocks: a block in a closed loop with a high-contrast visual cue, a block in a closed loop with wind and a block in a closed loop with 'wind' but with the air turned off (meaning that the nozzle moved around the fly as normal but no air was flowing). We found that the bump had a consistent offset relative to the virtual environment in the first two blocks (as expected) but it drifted relative to the environment in the third block (as we would expect from a fly walking in darkness); in other words, HD encoding accuracy was high in the first two blocks and low in the third block. This result confirmed that the visual image of the nozzle did not act as an effective HD cue.

Wind direction was controlled using Python and Arduino. The stepper motor was controlled using a controller board (X-NUCLEO-IHM02A1, STMicroelectronics) and Arduino UNO. Arduino code allowed us to specify the location of the air nozzle in the next time step and a custom Python code communicated with the Arduino through serial port to generate the pattern of nozzle movements. When the wind was in closed loop with the fly's rotation, the Python code obtained the current HD of the fly through FicTrac (as described above) and generated the command to move the air nozzle in the appropriate location, such that the allocentric wind direction remained constant, from the fly's perspective.

The wind speed was  $0.2 \text{ m s}^{-1}$ , measured with a hot-wire anemometer (A004, Kanomax) at the fly's location. Air flow was regulated through a mass flow controller (Aalborg, GFC17A-VAL6-CO). Before each experiment involving wind stimuli, we confirmed that the air nozzle was accurately pointing at the fly's antennae by observing the movement of the arista in response to wind using a camera (BFLY-PGE-31S4M, FLIR) equipped with a high-magnification lens (InfinitiStix lens 44 mm,  $\times 3.00$ , Infinity Photo-Optical).

### Stimulus protocols

For Figs. 2 and 3 and Extended Data Figs. 2 and 8, flies were in a closed loop for 20 min with a stimulus that switched across three contrast levels in 200-s blocks. The bright cue was a bright stripe (brightness level 15/15) against a black background (0/15). The dim cue was a dim stripe (1/15) against a black background (0/15). In the no-cue condition, the panorama displayed a black background (0/15). The block sequence was drawn randomly for each fly. Figure 2 uses data from all blocks, while Fig. 3 focuses only on each fly's first block with the dim cue, because that was the block where HD encoding accuracy was most variable across individuals. Extended Data Fig. 2 uses data from the first block of each type. Extended Data Fig. 8 uses data from the transitions between darkness and the bright cue.

For Fig. 4 and Extended Data Fig. 3, flies were first in a closed loop for 10 min with a single cue: a bright stripe (15/15) against a dark background (0/15) or wind. Then, flies were in a closed loop for 5 min with the cue they did not receive in the first block. Next, the flies were in a closed loop with both cues for 45 min. Every 5 min, one of the cues shifted, with alternating visual shifts and wind shifts. If the fly had been presented with a visual cue as the first stimulus, then the visual cue was

the first cue to jump and vice versa. Cue shifts were  $+120^\circ$  or  $-120^\circ$ , with the direction determined randomly for each shift. Because of the nature of our stimulus delivery, the visual cue shift occurred essentially instantaneously but the wind shift was limited by the stepper motor speed, taking  $\sim 0.4 \text{ s}$  to rotate to its new location.

For Fig. 5 and Extended Data Fig. 4, flies were first in a closed loop for 10 min with a single cue: a bright stripe (15/15) against a black background (0/15) or wind. Then, flies were in closed loop for 10 min with the cue they did not receive in the first block. The block order (visual then wind or wind then visual) was drawn randomly for each fly. Next, flies were in a closed loop with both cues presented simultaneously for 800 s. In the fourth block, flies were again in a closed loop with the initial single cue, this time for 5 min. Finally, flies were in a closed loop with the second single cue, also for 5 min.

For Figs. 6 and 7 and Extended Data Fig. 6, flies were first in a closed loop with a bright stripe (15/15) for 15–20 min to allow the compass system to stabilize; data from this stabilization epoch are not shown in the figures but instead used for Extended Data Figs. 5 and 7 (top). Next, flies were in a closed loop with a bright stripe for 20 min. For the first 200 s, the gain between the fly's movement and the cue movement was set to 1 (normal gain); for the next 800 s, it was set to  $-1$  (inverted gain); for the final 200 s, it was reverted back to 1. Finally, flies walked in darkness for 200 s.

For Extended Data Figs. 5 and 7 (top), flies were in closed loop with a bright stripe (15/15) against a black background (0/15) for 15–20 min.

For Extended Data Fig. 7 (bottom), these panels show responses to an open-loop presentation of a rotating visual cue. The cue was a bright stripe (15/15) against a black background (0/15). Each fly received 14 60-s blocks, with the cue rotating at  $30^\circ$ ,  $60^\circ$ ,  $90^\circ$ ,  $120^\circ$ ,  $150^\circ$ ,  $180^\circ$  or  $200^\circ$  per second (two blocks per stimulus speed, with speed order randomized).

**Randomization and blinding.** The experimenter was not blind to the experimental conditions assigned to each fly. Blinding is only potentially relevant to experiments where different flies experienced different experimental treatments (for example, different genotypes or different drug treatments). However, there was only one experiment where different flies experienced different experimental treatments, namely the cases where the order of the different brightness blocks was randomized (Figs. 2 and 3 and Extended Data Figs. 2 and 8). Here, the experimenter could not be blinded to the experimental condition because the stimulus apparatus delivered the stimulus in a manner that was visible to the experimenter.

### Data analysis

Data analysis was performed using Matlab R2021a (MathWorks, RRID:SCR\_001622), Python 3.7 and 3.9 ([www.python.org/](http://www.python.org/), RRID:SCR\_008394), Stan<sup>66</sup> ([mc-stan.org/](http://mc-stan.org/), RRID:SCR\_018459), PyStan 2.19.0 ([pystan2.readthedocs.io/en/latest/](http://pystan2.readthedocs.io/en/latest/)), R 4.1.3 ([www.r-project.org/](http://www.r-project.org/), RRID:SCR\_001905) and RStudio 2022.02.0 ([www.rstudio.com/](http://www.rstudio.com/), RRID:SCR\_000432). Tukey tests and Pearson correlations assumed that data were distributed normally but this was not formally tested. No flies were excluded from analyses.

**Preprocessing imaging data.** The NoRMCorre algorithm<sup>67</sup> (<https://github.com/flaironinstitute/NoRMCorre>) was used to perform rigid motion correction in the  $x$ ,  $y$  and  $z$  dimensions. Then, the maximum  $z$ -projection was used to select a mask around the PB and define our region of interest (ROI). The PB was divided along its horizontal axis into 25–40 ROIs (Extended Data Fig. 1). The change in fluorescence ( $\Delta F/F$ ) was computed for each ROI, with the baseline fluorescence  $F$  defined as the bottom tenth percentile of fluorescence values for the trial for that ROI.

**Analysis of locomotion data.** The position of the ball in all three axes was computed by FicTrac at 50 Hz. This was used to infer the position of the fly in fictive two-dimensional space. Position data were unwrapped

and then smoothed using locally weighted scatter plot smoothing (LOWESS smoothing). The velocity was computed in all three axes and smoothed again with the LOWESS method. The smoothed data were then downsampled to match the imaging volume rate.

**Bump width and amplitude.** A von Mises function of the form  $f(x) = \alpha \exp(\kappa \cos(x - \mu)) + c$  was fit to each time point of our  $\Delta F/F$  signal with Matlab's 'fit' function using the nonlinear least squares method and the trust-region algorithm for optimization. The estimated value of  $\mu$  represents the bump's position. The fit coefficients were then used to compute the bump width and amplitude:

$$\text{bump width} = 2 \left| \cos^{-1} \left[ \frac{1}{\kappa} \log \left( \frac{1}{2} (\exp(\kappa) + \exp(-\kappa)) \right) \right] \right|$$

$$\text{bump amplitude} = a (\exp(\kappa) - \exp(-\kappa))$$

A goodness-of-fit metric (adjusted  $R^2$ ) was obtained for each time point and time points were discarded from group analyses if the adjusted  $R^2$  was below 0.5. Extended Data Figure 1 shows examples of a good fit and a poor fit.

**HD.** HD was taken as the time integral of the fly's rotational velocity on the spherical treadmill, relative to the position with the cue directly in front of the fly ( $0^\circ$ ). In the two-cue environment in Fig. 5, the wind tube was aligned with the visual cue; thus, HD =  $0^\circ$  corresponded to the position where both the wind tube and the visual cue were directly in front of the fly. The bump position rotates clockwise in the ellipsoid body (EB; imaged from the posterior side of the head) as HD rotates counterclockwise; therefore, to account for this directionality when we plotted HD over time (for example, in Fig. 1f and elsewhere), we always plotted ( $-HD$ ) to make it easier to visualize the correspondence between bump position and HD.

**Offset and HD encoding accuracy.** The offset of the bump relative to HD was computed as the circular distance between bump position and HD. Bump position rotates clockwise in the EB (imaged from the posterior side of the head) as HD rotates counterclockwise; therefore, to account for this directionality, we used ( $-HD$ ) rather than HD:

$$\text{offset} = \text{bump position} - (-HD)$$

In a single-cue environment, we defined HD =  $0^\circ$  as the value of HD that places the cue directly in front of the fly; thus, an offset of  $+90^\circ$  means that the bump is positioned at  $+90^\circ$  (on the right-hand edge of the EB) when the fly is facing the cue. To compute HD encoding accuracy, each offset value was treated as a unit vector and the vector strength of these values was calculated. HD encoding accuracy was computed only over the time points when the fly was moving; that is, moments of immobility were excluded.

**Bump preference index and behavioral preference index.** In Fig. 4 and Extended Data Fig. 3, we computed a bump preference index for each cue shift. Here, we first obtained the mean value of the visual cue and wind offset before the cue shift (calculated over the 2-min window preceding each shift) and following the cue shift (calculated over a 2-min window starting 30 s after the cue shift because we found that it took about 30 s for the offset to stabilize after a cue shift). We then calculated the change in visual cue offset and wind offset by taking the difference between the postshift and preshift values. We computed the bump preference index as follows:

$$\text{bump preference index} = \frac{|\text{change in visual cue offset}| - |\text{change in wind offset}|}{|\text{change in visual cue offset}| + |\text{change in wind offset}|}$$

The stickiness index was obtained in the same way, except that the offset here was computed relative to the shifted cue and the nonshifted

cue. Here, a value of +1 means that the bump 'sticks' with the nonshifted cue, whereas a value of  $-1$  means that the bump follows the shifted cue:

$$\text{stickiness index} = \frac{|\text{change in shifted\_cue offset}| - |\text{change in nonshifted\_cue offset}|}{|\text{change in shifted\_cue offset}| + |\text{change in nonshifted\_cue offset}|}$$

The behavioral preference index was computed in the same way, except that the offset here was the angular difference between the fly's HD and the cue position:

$$\text{behavioral preference index} = \frac{|\text{change in visual cue offset}| - |\text{change in wind offset}|}{|\text{change in visual cue offset}| + |\text{change in wind offset}|}$$

**Offset changes during configurational learning.** In Fig. 5, we defined the amount of conflict in the two-cue environment as follows:

$$\theta_n = \text{offset with both cues} - \text{initial single cue offset}$$

The resulting offset change was determined as follows:

$$\phi_n = \text{final single cue offset} - \text{initial single cue offset}$$

In Fig. 5d, to fit the relationship between the conflict  $\theta_n$  and the amount of remapping  $\phi_n$ , we used a probabilistic model:

$$\mu_n = a \cdot \theta_n + b$$

$$\phi_n \sim \text{von Mises}(\mu_n, \kappa)$$

In other words, we assumed that  $\phi_n$  is generated by adding a noise (distributed according to the von Mises distribution) to an intermediate variable  $\mu_n$ , which is assumed to be a linear function of  $\theta_n$  with slope  $a$  and offset  $b$ . The magnitude of the noise is characterized by the concentration parameter  $\kappa$  of the von Mises distribution. We obtained estimates of parameters  $a$ ,  $b$  and  $\kappa$  by performing Bayesian analysis with the following prior distributions over the parameters:

$$a \sim N(0, 1)$$

$$b \sim \text{uniform}(-\pi, \pi)$$

$$\kappa \sim \text{inverse gamma}(1.91, 6.47)$$

We chose a prior over parameter  $a$  on the basis of prior knowledge that a slope with large magnitude is highly unlikely. We chose a uniform prior over the offset  $b$ . We chose an inverse gamma distribution over the concentration parameter  $\kappa$  to suppress large  $\kappa$  values (all points lie perfectly on a straight line) and to strongly suppress small  $\kappa$  values (large noise)<sup>68</sup>. The parameters of the inverse gamma were chosen such that <2% of  $\kappa$  fell below 1 or above 50. We confirmed that these prior choices are reasonable according to prior predictive simulations.

We estimated the three parameters by obtaining samples from the posterior using the Hamiltonian Monte Carlo implemented in Stan 2.19.0 (ref. 66) (four chains with 1,000 samples each after 1,000 samples of warm-up, no thinning) accessed using PyStan2. Diagnostics of the Hamiltonian Monte Carlo did not show any issues with the fit, examination of the posterior distribution using the pair plots indicated no signs of multimodality and the posterior retrodictive checks showed no obvious discrepancies between the model and the data<sup>69</sup>.

**Remapping in inverted gain.** In Fig. 7 and Extended Data Fig. 6, the remapping index is the difference between the HD encoding accuracy with respect to the fly's HD and the HD encoding accuracy with respect to the visual cue position, excluding time points where the fly was immobile. Specifically,



remapping index

= vector strength (offset<sub>bump-HD</sub>) – vector strength(offset<sub>bump-visual</sub>)

Data points when the fly was immobile were discarded from the remapping index calculation. This index is designed to be robust to the spontaneous alterations in the fly's behavior. For example, if the fly were to walk perfectly straight for some time and if its EPG activity bump did not move during this time, then vector strength(offset<sub>bump-HD</sub>) and vector strength(offset<sub>bump-visual</sub>) would both be 1; taking the difference between these two vector strength values yields a remapping index of 0, which captures the fact that the bump would be equally well aligned with the fly's self-motion and the visual cue.

## Network model

Our network model follows the structure of a previously published ring attractor model<sup>22,24</sup>, with several key modifications. First, we systematically varied the amplitude of sensory (ER) inputs to the ring attractor. Second, we gave the attractor network two independent streams of sensory input, with distinct ER → EPG weights. Third, we updated two learning parameters ( $w_{\max}$  and  $g_0$ ) to promote stable outcomes after learning, while also maintaining the sensitivity of the network to all available cues. Details on these changes are provided below, along with a summary of the model infrastructure.

In this model, the firing rate dynamics of the  $N$  simulated EPG neurons are given by

$$\tau \frac{df_n}{dt} = -f_n + [\alpha f_n + D(f_{n-1} + f_{n+1}) + \frac{v(t)}{v_{\text{rel}}} \frac{1}{2}(f_{n+1} - f_n) - \beta \sum_l f_l + I_{n,1}(t) + I_{n,2}(t) + 1]_+$$

where  $f_n$  is the firing rate of neuron  $n$ ,  $\tau$  is the network's time constant,  $[\cdot]_+$  is the linear-rectifying function that returns  $[a]_+ = a$  if  $a > 0$  and  $[a]_+ = 0$  otherwise,  $\alpha$  and  $D$  control the local self-excitation,  $v(t)$  is the fly's current angular velocity,  $v_{\text{rel}}$  is a parameter that controls how this angular velocity impacts EPG activity,  $\beta$  controls the amount of global inhibition and  $I_{n,k}(t)$  is the external input from cue  $k \in \{1, 2\}$  to EPG neuron  $n$ .

Inhibitory inputs indicating cue positions,  $I_k = (I_{1,k}, \dots, I_{N,k})^T$ , are formed by

$$I_k(t) = -W_k(t)g_k(t)$$

where  $g_k = (g_{1,k}, \dots, g_{M,k})$  is the vector defining the activity of the MER neurons associated with cue  $k$  and  $W_k$  is the  $N \times M$  weight matrix whose  $nm$ th element  $W_{k,nm}$  specifies the synaptic weight from the  $m$ th ER neuron to the  $n$ th EPG neuron. Weights are non-negative; thus,  $I_k \leq 0$ . These synaptic weights are continuously updated according to a postsynaptically gated learning rule

$$\frac{dW_{k,nm}}{dt} = \eta |v(t)| f_n (w_{\max}(1 - g_{k,m}(t)/g_0) - W_{k,nm}(t))$$

where  $\eta(|v(t)|)$  is the learning rate (which depends on the absolute value of the fly's rotational velocity  $v(t)$ ) and  $w_{\max}$  and  $g_0$  are parameters that control the learning dynamics of the different parts of the learning rule. This learning rule can be rewritten as follows:

$$\frac{dW_{k,nm}}{dt} = \eta |v(t)| f_n (w_{\max} - W_{k,nm}(t)) - \eta(|v(t)|) f_n w_{\max} g_{k,m}(t)/g_0$$

where the first term represents nonassociative LTP and the second term represents associative LTD. Here, nonassociative LTP depends on postsynaptic activity but not presynaptic activity; by contrast, associative LTD depends on both presynaptic and postsynaptic activity. Activity of the  $k$ th ER neuron population is given by

$$g_{k,m}(t) = \varepsilon_{\text{ER}}(t) + A_k \exp[\kappa_k \cos(\theta(t) - \theta_m) - 1]$$

where the upper bound of the uniformly distributed baseline noise  $\varepsilon_{\text{ER}}(t) \sim \text{uniform}(0, b_{\text{ER}} \sum_i f_i^{\text{ss}})$  is the summed activity of the EPG neurons in a steady state without external inputs,  $I_{n,1}(t) = I_{n,2}(t) = 0$ , scaled by a constant factor  $b_{\text{ER}}$ ,  $A_k$  is the amplitude of the population activity profile (ER amplitude),  $\kappa_k$  is the precision (inverse width) of the population activity profile and  $\theta_m$  is the preferred HD of the  $m$ th ER neuron.

We simulated the fly's HD sequence by drawing a sequence of HD displacements,  $du(t) \sim N(0, \sigma_u^2 dt)$ , from a zero-mean Gaussian with variance  $\sigma_u^2 dt$ , which we then turned into a sequence of angular velocities  $du(t)/dt$ . To ensure a smooth angular velocity sequence, we then applied a running window average of the  $du(t)/dt$  sequence (centered window, size 2.5 s) to generate the actual angular velocity sequence  $v_{\text{true}}(t)$ . This sequence was integrated across time to yield the HD sequence  $\theta_{\text{true}}(t)$ . The activity of the EPG neurons  $f_n$  was driven by a noisy version of this angular velocity,  $v(t) = v_{\text{true}}(t) + \varepsilon_{\text{AV}}(t)$ , where  $\varepsilon_{\text{AV}}(t)$  is a Gaussian white noise process with s.d.  $\sigma_{\text{AV}}$  that has been smoothed with a running window average (centered window, size 0.04 s).

In all network simulations, we either simulated individual trials or simulated a set of trials and then averaged across those trials. In each trial, the EPG activity and synaptic weight dynamics were simulated by Euler integration and the ER activity and the fly's HD sequence were generated in time steps of  $\Delta t = 2.5$  ms. The EPG activity was initialized by a cosine profile and then simulating the network for 20 s in the absence of both angular velocity and external inputs,  $I_{n,1}(t) = I_{n,2}(t) = 0$ , such that the EPG activity reached an approximate steady state  $f^{\text{ss}}$ . The synaptic weights were initialized randomly by drawing their elements from a uniform distribution over  $[0, 1]$  and subsequently normalized to reach an initial matrix Frobenius norm of  $\|W_k\|_F = 1.5$ . This was followed by a burn-in period (analysis-dependent duration; discussed below) such that synaptic weights could approximately reach a steady state. What followed was dependent on the specific analysis in question and is described in detail below. Unless otherwise mentioned, we simulated a network with  $N = M = 32$  EPG and ER neurons (per external input) and used the following parameters: EPG network time constant  $\tau = 50$  ms, local excitation  $\alpha = -8.93$  and  $D = 5.19$ , global inhibition  $\beta = 0.11$ , angular velocity scaling  $v_{\text{rel}} = 3.64$ , synaptic weight learning rate  $\eta = 0.34$ , learning parameters  $w_{\max} = 1/17$  and  $g_0 = 1$ , input activity baseline noise factor  $b_{\text{ER}} = 0.45$ , fly motion noise  $\sigma_u = 8$  radians per second and angular velocity noise  $\sigma_{\text{AV}} = 1$  radians per second. Note that these values of  $w_{\max}$  and  $g_0$  are different from those used in previous studies; we found that it was necessary to change these values to balance stability and flexibility during learning. In particular, if learning led to weight magnitudes that were too large, the network model ignored angular velocity inputs; on the other hand, if learning led to weight magnitudes that were too small, the network model never established a stable ER → EPG mapping. To simulate the ER activity  $g_k(t)$ , we specified a scale factor  $\tilde{A}_k$  (analysis-dependent scale factor; discussed below) that was multiplied by the amplitude of the steady-state EPG activity bump  $f^{\text{ss}}$  (1.062) to obtain the ER amplitude  $A_k \approx 1.062 \tilde{A}_k$ . For all simulations, we set the width of the ER activity bump to be  $w_k = 0.8$  for all  $k$ , which was then converted to the precision by  $\kappa_k = \log(2)/(1 - \cos(w_k/2))$ . For simulations in which only a single cue was present, we fixed  $I_{2,n} = 0$  for all  $n$ . In the list of parameter values above, all reported times are simulation times. When plotting time courses, we converted them to experiment time by assuming that 24 s of experiment time corresponded to 1 s of simulation time. We assessed HD encoding accuracy by first computing the circular variance between HD  $\theta_{\text{true}}$  and the position of the EPG activity bump  $\arg \max_f f_i$  on  $[-\pi, \pi]$  over a causal 8-s window. The HD encoding accuracy was defined as  $1 - \text{circular variance}$ . We computed bump width and amplitude by finding the full width at half maximum and the



difference between the peak and the trough, respectively, of the EPG activity bump. We computed the weight matrix notch depth by first smoothing the weight matrix  $W_k$  with a Gaussian filter with s.d.  $\sigma_{\text{smooth}} = 2$  and then finding the difference between the maximum and the minimum matrix elements.

**Impact of cue intensity on bump parameters and across-individual variability.** In Figs. 2h,i and 3d,f, we simulated a network with a single external input with different cue intensities (41 values of  $\tilde{A}_k$  from 0 to 2 in steps of 0.05). We assessed the bump parameters in a 30-s period after a burn-in period of 120 s. The bump parameters shown in Fig. 2h are averages across 100 simulated trials for each cue intensity. To simulate across-individual variability, we generated individual variations in ER amplitude in response to a given cue. We, thus, plotted the bump width (Fig. 3d) and bump amplitude (Fig. 3f) over HD encoding accuracy for different cue intensity (11 values of  $\tilde{A}_k$  from 0.5 to 1 in steps of 0.05) and five different simulated trials per cue intensity.

**Impact of individual variations in experienced cue intensity on cue weighting.** In Fig. 4f,g, we simulated a network with two external inputs. To introduce variations and differences in experienced cue intensity, we varied the ER amplitude (11 values of  $\tilde{A}_k$  from 0 to 1 in steps of 0.05) of one ER population while holding the ER amplitude of the other ER population constant (ER amplitude scale set to 0.75) and vice versa, producing 22 ER amplitude pairs. ER amplitude was set to 0 when the associated cue was off. Each simulated trial began with a 30-s burn-in period during which both cues were off. The burn-in period was followed by two single-cue blocks where each cue was turned on for 25 s one at a time. We assessed the bump width, amplitude and HD encoding accuracy during the second half of each single-cue block. After the single-cue blocks, both cues were on for a total of 262.5 s. Every 12.5 s, one of the cues shifted, with the shifts alternating between visual cue and wind cue. Thus, each simulated trial comprised 20 cue shifts with ten shifts per modality. Cue shifts were  $+120^\circ$  or  $-120^\circ$ , with the direction determined randomly for each shift. If visual cue was on in the first single-cue block, then visual cue was the first cue to shift and vice versa. The order was balanced across trials such that visual cue shifted first in half of the trials. For each cue shift, we computed a bump preference index (discussed above) using simulated data from a 5-s period before and a 5-s period after the cue shift. We simulated three trials for each combination of ER amplitude pair and cue order, producing 132 trials in total. The scatter points in Fig. 4f and vertical lines in Fig. 4g each represent a single simulated trial. Each scatter point in Fig. 4g is an average across ten cue shifts of the same modality from the same trial.

**Temporal evolution of bump parameters during cue combination.** In Fig. 5f,g, we simulated a network with two equally strong external inputs. When a cue was on, its associated  $\tilde{A}_k$  was set to 2; when a cue was off, its associated  $\tilde{A}_k$  was set to 0. Each simulated trial began with a burn-in period of 30 s during which both cues were off. The burn-in period was followed by two 25-s single-cue blocks; cue 1 was on while cue 2 was off in the first single-cue block and vice versa in the second single-cue block. This was followed by a two-cue block during which both cues were on for 32.5 s. The two-cue block was followed by two 12.5-s single-cue blocks, again with cue 1 on in the first and cue 2 on in the second. We visualized the temporal evolution of the bump parameters after the burn-in period (Fig. 5f). The bump parameters shown in Fig. 5f are averages across 100 simulated trials.

**Impact of variation in ER amplitude on remapping during inverted gain.** In Fig. 7g,h, to model across-individual variability, we simulated a network with a single external input with five values of  $\tilde{A}_k$  ranging from 1 to 2 in steps of 0.25. Each simulated trial began with a burn-in period of 60 s during which the gain between the simulated fly's

movement and the cue movement was set to 1. Following the burn-in period, the gain remained at 1 for another 8 s (normal gain), after which it was set to  $-1$  for 32 s (inverted gain). We visualized the temporal evolution of the bump parameters after the burn-in period (Fig. 7g). The bump parameters shown in Fig. 7g are averages across 100 simulated trials for each ER amplitude.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All data are available from the DANDI repository (<https://dandiarchive.org/dandiset/000289>).

## Code availability

Data analysis code is available via Zenodo at <https://doi.org/10.5281/zenodo.15733533> (ref. 70). The model code is also available via Zenodo at <https://doi.org/10.5281/zenodo.13270653> (ref. 71). Simulated data to recreate figures with the model code are available via figshare at <https://doi.org/10.6084/m9.figshare.26510239> (ref. 72).

## References

- Jenett, A. et al. A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep.* **2**, 991–1001 (2012).
- Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
- Moore, R. J. D. et al. FicTrac: a visual method for tracking spherical motion and generating fictive animal paths. *J. Neurosci. Methods* **225**, 106–119 (2014).
- Reiser, M. B. & Dickinson, M. H. A modular display system for insect behavioral neuroscience. *J. Neurosci. Methods* **167**, 127–139 (2008).
- Handler, A. et al. Distinct dopamine receptor pathways underlie the temporal sensitivity of associative learning. *Cell* **178**, 60–75 (2019).
- Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* <https://doi.org/10.18637/jss.v076.i01> (2017).
- Pneumatikakis, E. A. & Giovannucci, A. NoRMCorr: an online algorithm for piecewise rigid motion correction of calcium imaging data. *J. Neurosci. Methods* **291**, 83–94 (2017).
- Betancourt, M. Containment prior modeling case study. *GitHub* [https://github.com/betanalphabet/knitr\\_case\\_studies/tree/master/some\\_containment\\_prior\\_models](https://github.com/betanalphabet/knitr_case_studies/tree/master/some_containment_prior_models) (2021).
- Betancourt, M. Towards a principled Bayesian workflow (PyStan). *GitHub* [https://github.com/betanalphabet/jupyter\\_case\\_studies/tree/master/principled\\_bayesian\\_workflow](https://github.com/betanalphabet/jupyter_case_studies/tree/master/principled_bayesian_workflow) (2018).
- Basnak, M. Analysis code for Basnak et al. Nature Neuroscience 2025. *Zenodo* <https://doi.org/10.5281/zenodo.15733533> (2025).
- Drugowitsch, J. Multimodal HD cue integration sims: code as used in manuscript. *Zenodo* <https://doi.org/10.5281/zenodo.13270653> (2024).
- Drugowitsch, J., Chen, A. & Kutschireiter, A. Network simulation data. *figshare* <https://doi.org/10.6084/m9.figshare.26510239> (2024).

## Acknowledgements

We thank R.I.W. lab members for feedback on the manuscript. M. Dickinson provided help with the spherical treadmill apparatus. This work was supported by grants from the National Institutes of Health (R34NS123819 to J.D. and R.I.W., U19NS104655 to R.I.W., U19NS118246 to J.D. and P40OD018537 to the Bloomington *Drosophila* Stock Center). M.A.B. was supported by a Stuart H.Q. and Victoria

Quan fellowship from Harvard Medical School. A.K. was supported by Early Postdoc.Mobility (P2ZHP2\_184213) and Postdoc.Mobility (P400PB\_199242) fellowships from the Swiss National Science Foundation. T.S.O. was supported by a distinguished scholar award from the Warren Alpert Foundation. A.C. was supported by a National Science Foundation graduate research fellowship (DGE2140743). R.I.W. is a Howard Hughes Medical Institute investigator.

### Author contributions

M.A.B. designed and performed experiments and data analyses, with advice from A.K., R.I.W. and J.D. T.S.O. and P.G. designed and built the virtual environment combining visual and mechanosensory stimuli, with help from M.A.B. A.K. and A.C. performed computational modeling, with advice from J.D., M.A.B. and R.I.W.

### Competing interests

The authors declare no competing interests.

### Additional information

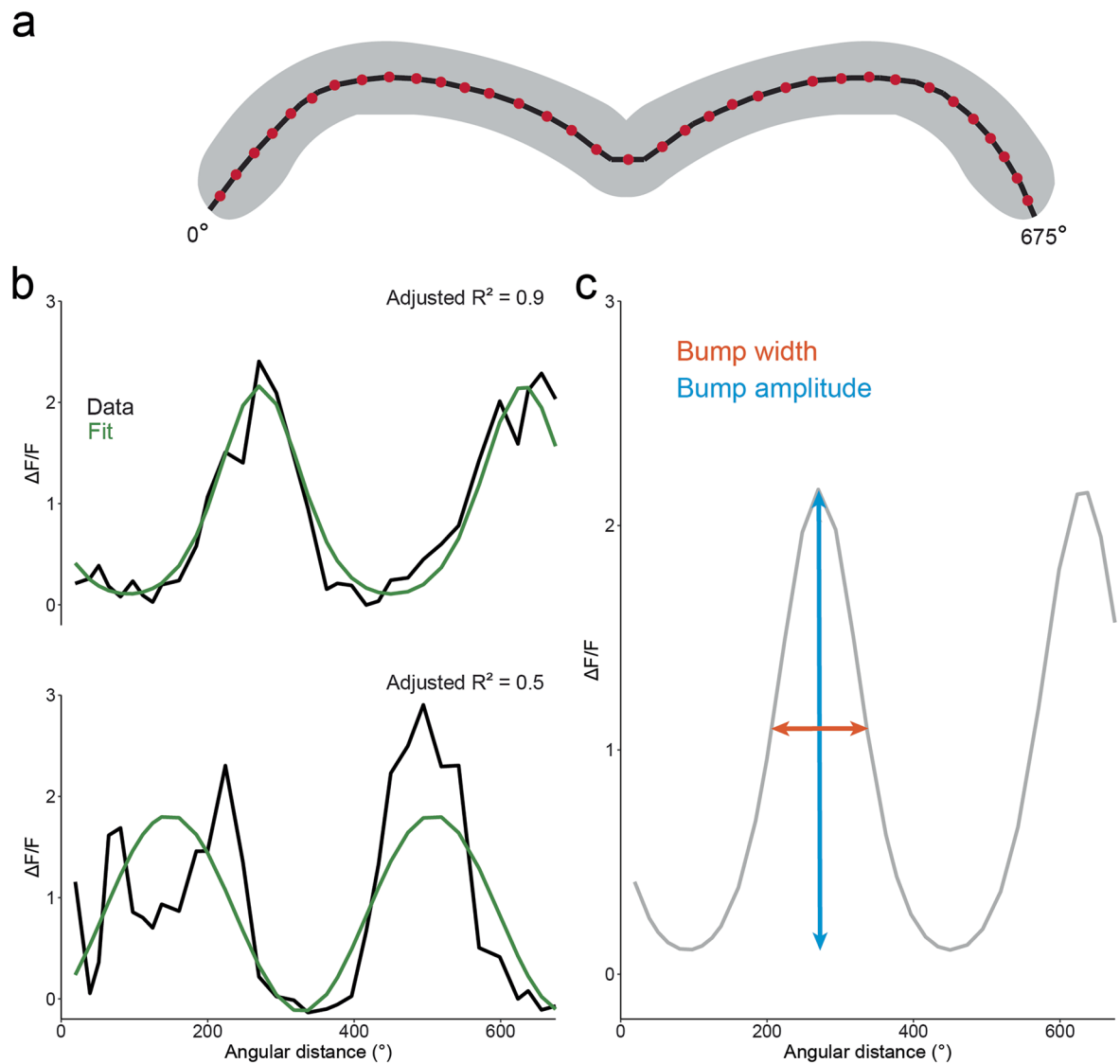
**Extended data** is available for this paper at <https://doi.org/10.1038/s41593-024-01823-z>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41593-024-01823-z>.

**Correspondence and requests for materials** should be addressed to Rachel I. Wilson.

**Peer review information** *Nature Neuroscience* thanks Stanley Heinze and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

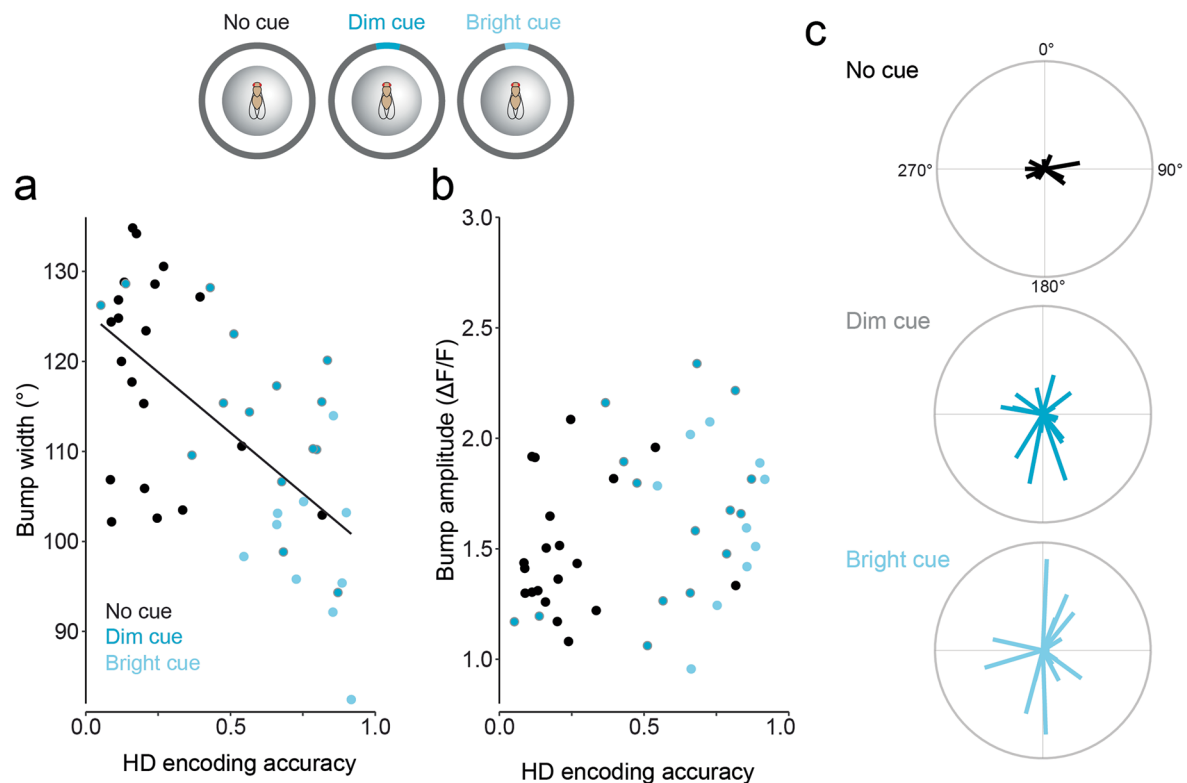
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



### Extended Data Fig. 1 | Fitting the bump of activity in the HD system.

**a.** Schematic illustrating procedure for dividing the PB into spatial bins. First, the maximum z-projection was used to draw a mask around the protocerebral bridge (gray). Then, the *drawpolyline* function in Matlab was used to trace a line (black) composed of many short segments (25–40 segments) running along the midline of the PB. Each segment defines a spatial bin (that is, a region of interest or ROI). Next, a custom algorithm was used to find the line bisecting each segment perpendicularly (a normal line), and then to find the shortest distance between each pixel in our PB mask and one of the normal lines. The signal from each pixel is then assigned to the ROI corresponding to that normal line, with the entire array of ROIs representing an angular distance of 675°, that is one complete HD representation (360°) plus 7/8 of a second representation (315°)<sup>41</sup>. Note that previous studies have divided the PB into 16 spatial bins<sup>37,38</sup>, but here we use many more spatial bins (up to 40). This is justified because the fluorescence emission

from each EPG neuron will tend to “spill out” from its home glomerulus, due to (1) the scattering of fluorescence emission into adjacent glomeruli, and (2) the presence of stray EPG axonal branches in adjacent glomeruli. As a result of this spillover, fluorescence signals are not perfectly compartmentalized within glomeruli. Thus, dividing the protocerebral bridge into more spatial bins should improve our ability to resolve the amplitude and shape of the bump of activity. Indeed, we found that this approach produced a more reliable measure of bump width, as indicated by higher variance explained (adjusted  $R^2$ ) when fitting linear models to bump width data (as in Figs. 3, 4, 7 and Extended Data Figs. 2, 5, 7). **b.**  $\Delta F/F$  and fitted function for two example time points. Shown are a case with very good goodness of fit (top), and a case with a borderline fit quality. In the latter case, adjusted  $R^2 = 0.5$ , which is our threshold for discarding data points from the analysis. **c.** Fit from the top example in panel b, showing the two parameters that we extract from the fit.

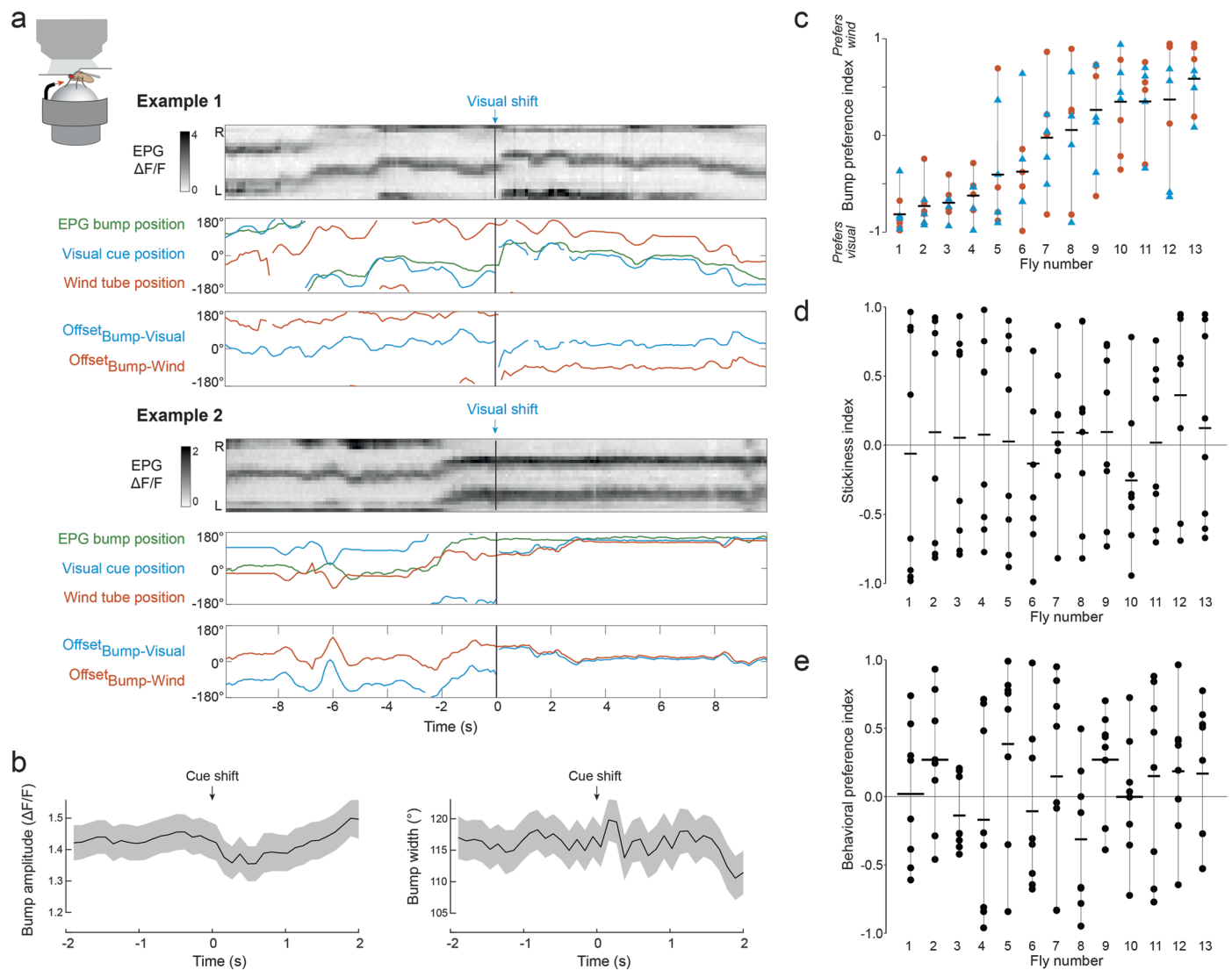


### Extended Data Fig. 2 | Individual variation in responses to cue brightness.

This figure provides additional detail related to the 15 flies in Fig. 3. **a.** Variations in HD encoding accuracy correlate with bump width, pooling data across all three cue conditions and all 15 flies ( $p < 0.001$ , Pearson's correlation). Gray points (dim cue) are reproduced from Fig. 3c. Black and blue points show data from the other two conditions (no cue and bright cue). **b.** There is a trend for variations in HD encoding accuracy to be correlated with bump amplitude, but the relationship

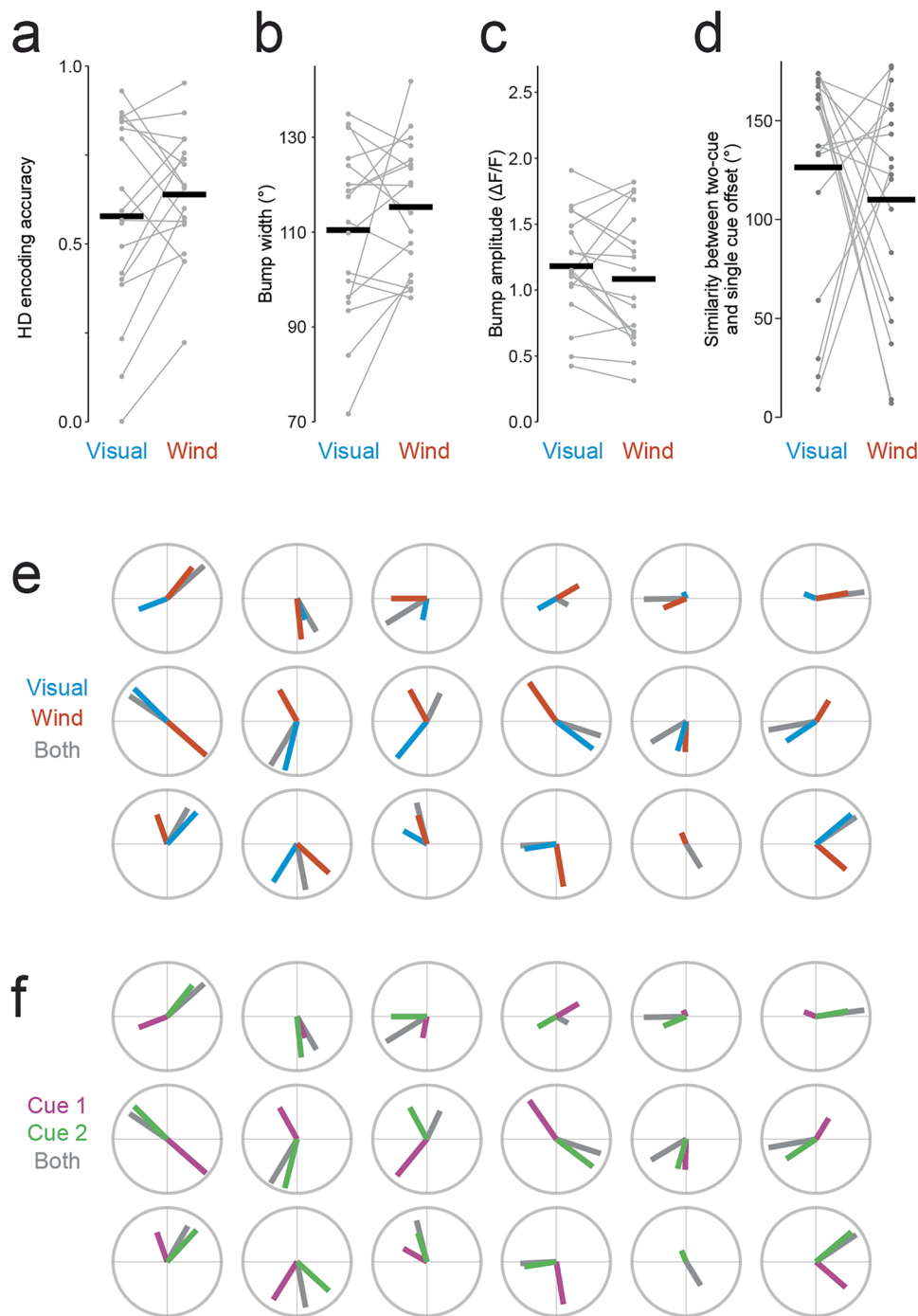
falls just short of significance ( $p = 0.06$ , Pearson's correlation). **c.** Mean head direction for each individual for all three cue conditions. The length of each line represents vector strength (that is, "consistency of behavioral orientation", or 1-circular variance). Note that consistency of behavioral orientation increases with increasing cue brightness. Different flies choose different goal directions, which is typical of navigation driven by the HD system<sup>32,33,42</sup>.





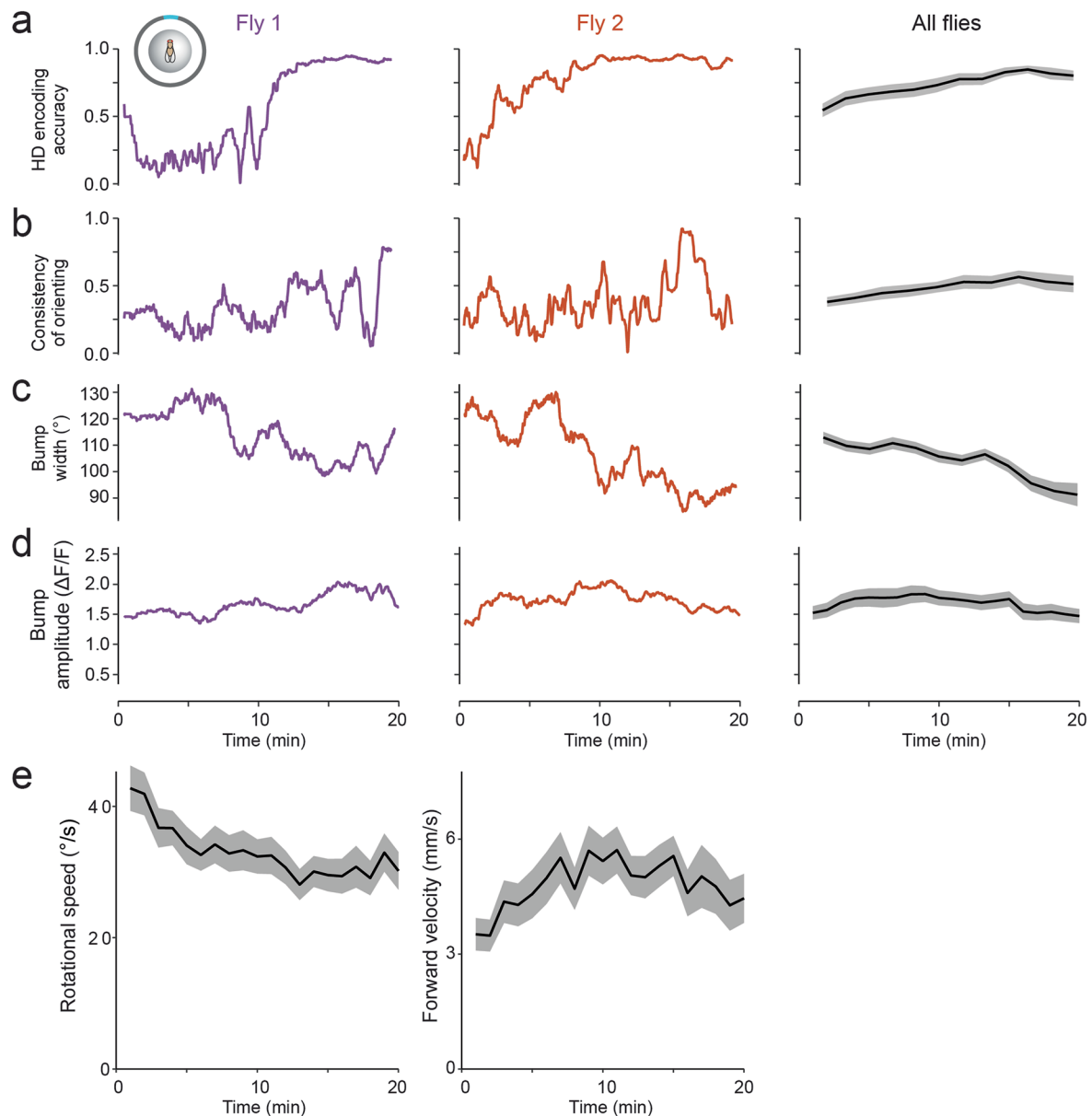
**Extended Data Fig. 3 | Effects of cue shifts.** This figure shows additional details regarding the 13 flies in Fig. 4. **a.** Responses to cue shifts. In both these examples, the visual cue shifted while the wind did not shift. In Example 1, the EPG bump follows the shift in the visual cue, so that the offset relative to that cue is restored. In Example 2, when the visual cue shifts, the bump does not shift, and so it retains its offset relative to the wind. **b.** Mean bump amplitude and mean bump width ( $\pm$  s.e.m.) around the time of cue shifts. Cue shifts produce no significant change in bump amplitude or width ( $p = 0.10$  for bump amplitude,  $p = 0.26$  for bump width, 2-sided Wilcoxon signed rank tests, comparing 1 sec before vs 1 sec after a shift). **c.** Bump preference index for each cue shift, for each fly, with flies sorted by mean bump preference index. Each thin vertical line connects the data for 4 wind shifts (orange) and 4 visual shifts (blue) for the same fly. Thick horizontal lines show mean values for each fly. This plot reproduces data from Fig. 4e.

**d.** Same but for the “stickiness index”. This index is a metric for how much the bump tends to follow the cue that does not shift (that is, how sticky the bump is). A value of +1 means that the bump sticks with the non-shifted cue, whereas a value of -1 means that the bump follows the shifted cue. There were no flies where the bump tended to prefer either the shifted cue or the non-shifted cue. Flies are sorted as in (c). **e.** Same but for the “behavioral preference index”. Here, -1 means the visual cue dominates the fly’s behavioral orientation, whereas +1 means the wind dominates. For example, if the wind shifts by  $-120^{\circ}$  and the fly then executes a  $-120^{\circ}$  turn, the index would be +1 for that trial. The mean behavioral preference index was near zero for all individuals, meaning they reoriented with equal frequency in response to visual cue shifts and wind shifts. This was true even in the flies with a strongly biased HD system (for example, Fly 1 or Fly 13).



**Extended Data Fig. 4 | Comparing one-cue environments and a two-cue environment.** This figure shows additional detail for the 18 flies in Fig. 5. **a–c.** There is no difference between the visual cue and the wind in terms of HD encoding accuracy ( $p = 0.28$ ; **a**), bump width ( $p = 0.27$ ; **b**), or bump amplitude ( $p = 0.20$ , paired 2-sided Wilcoxon tests; **c**). These plots show data from the two initial single-cue environments, prior to the onset of the two-cue environment. Thick horizontal lines are mean values. **d.** The offset during the two-cue block is equally similar to the initial offset with respect to the visual cue and the initial offset with respect to the wind ( $p = 0.52$ , paired Wilcoxon test). **e.** Offset data for both initial single-cue environments and the two-cue environment, for all 18 flies.

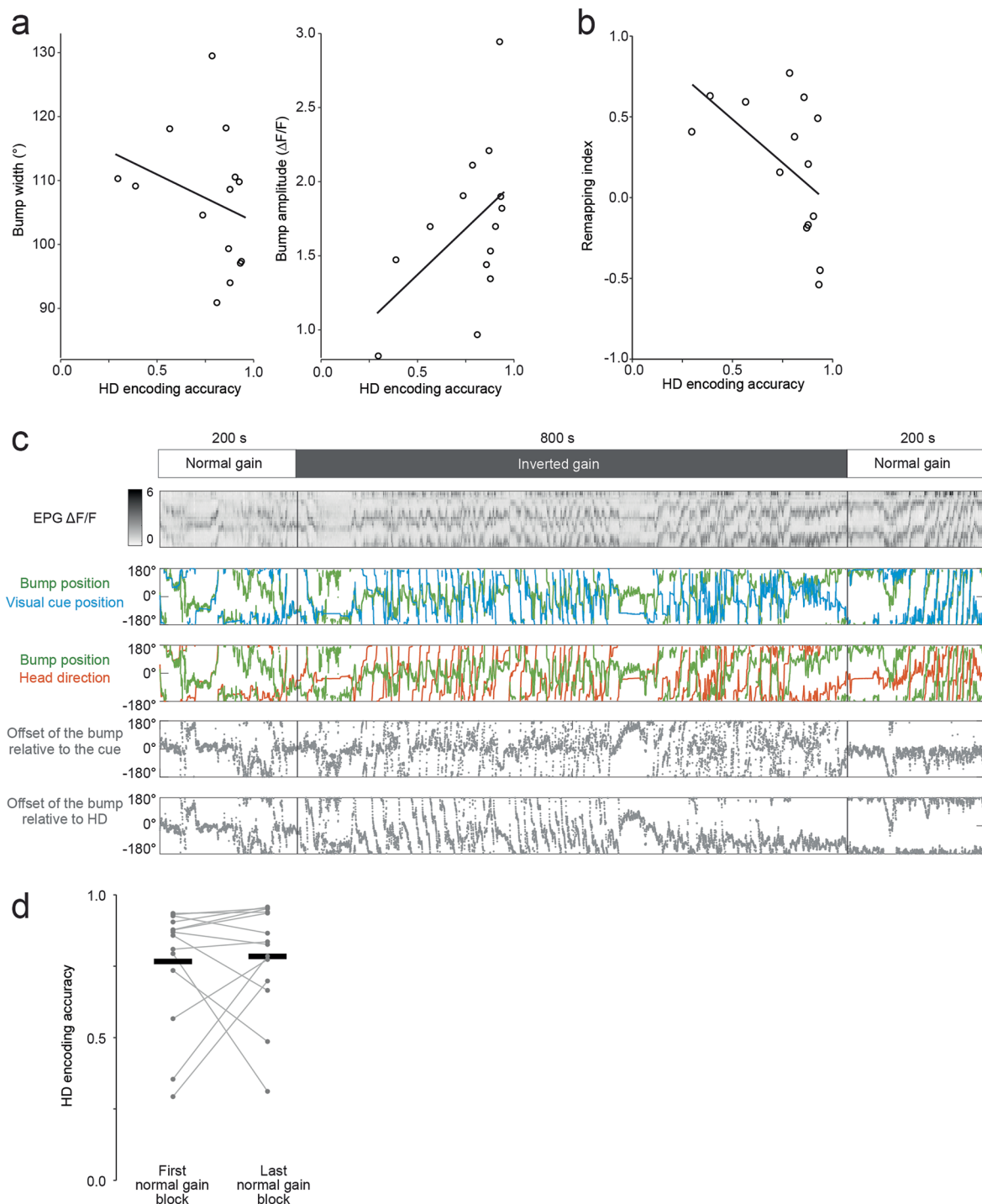
The angle of each line indicates the mean offset during that block; the length of each line indicates the vector strength of the offset (that is, HD encoding accuracy). Note that there is no relationship between the initial visual offset and the initial wind offset. Note also that there is no tendency for the visual cue or the wind to dominate the offset in the two-cue period. **f.** Same but now color-coded by cue order. Note that the offset in the two-cue environment is generally similar to the offset in the environment with cue 2. The exceptions to this rule are generally flies where the offset was not stable (for example, the fourth fly in the top row, or the fifth fly in the bottom row).



### Extended Data Fig. 5 | Acquiring familiarity with a virtual environment.

This figure shows data from 28 flies walking in the environment with the bright visual cue, beginning when they are first exposed to this environment. In **a-c**, The first two plots show data from two example flies and the third plot shows data for all flies (mean  $\pm$  s.e.m.). Each variable was smoothed with a 60-s moving window, after first excluding time points where the fly was immobile or goodness of fit was below threshold. **a**. Mean HD encoding accuracy increases over time. Over all flies, there is a significant effect of time in a linear mixed-effects model

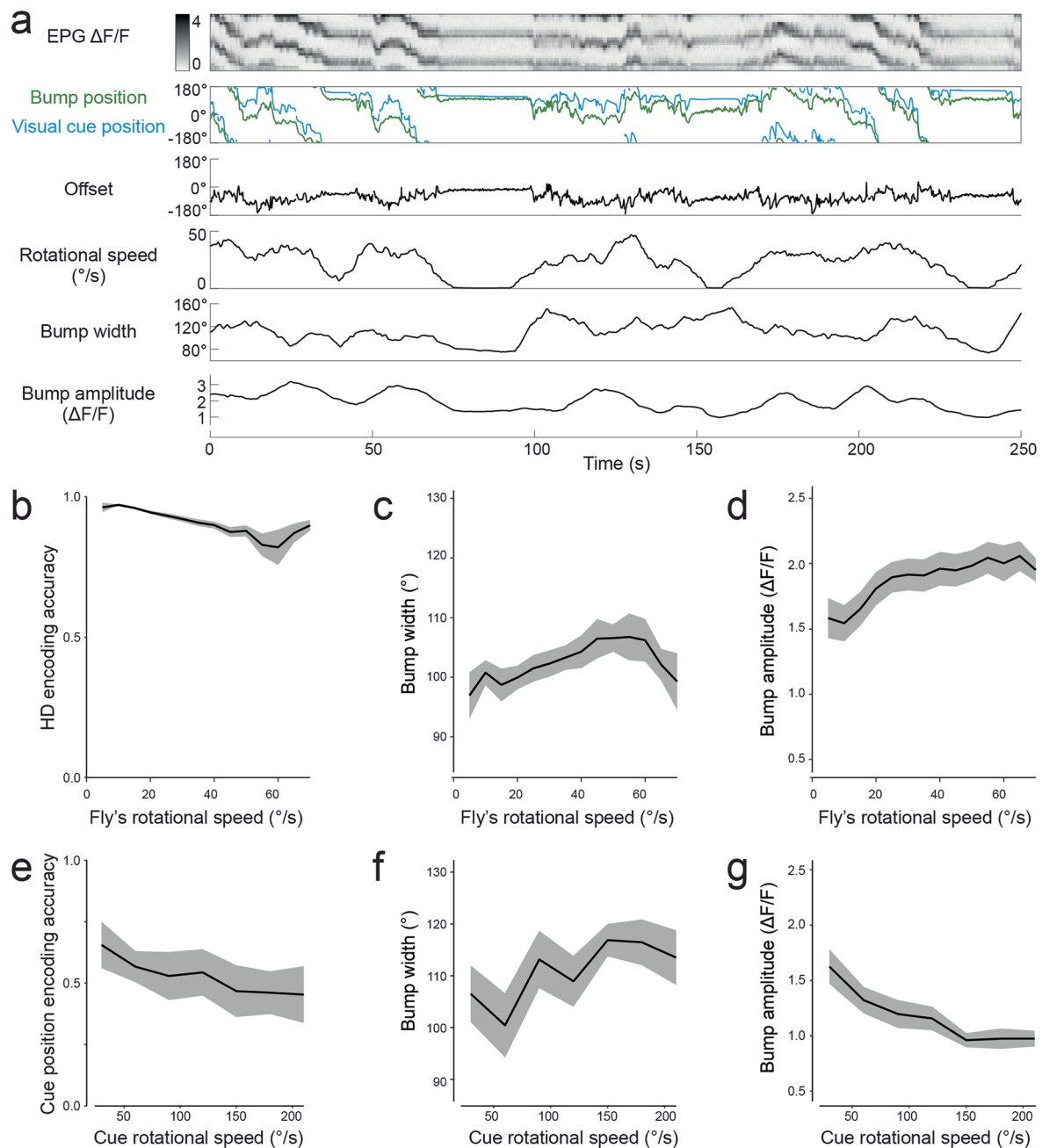
( $p < 0.001$ ). **b**. Mean consistency of behavioral orientation increases over time. Over all flies, there is a significant effect of time in a linear mixed-effects model ( $p < 0.001$ ). **c**. Mean bump width decreases over time. Over all flies, there is a significant effect of time in a linear mixed-effects model ( $p < 0.001$ ). **d**. Mean bump amplitude increases during the first 10 min. in the environment, and then decreases. Over all flies, there is a significant effect of time in a linear mixed-effects model ( $p = 0.03$ ). **e**. Mean rotational speed and forward velocity for the same flies over this time period ( $\pm$  s.e.m.).



**Extended Data Fig. 6 | Learning in inverted gain.** This figure shows additional details regarding the 14 flies in Fig. 7. **a.** In the initial normal-gain segment, bump width is significantly negatively correlated with HD encoding accuracy ( $p < 0.001$ , Pearson's correlation). There is a trend for bump amplitude to be positively correlated with HD encoding accuracy, but this relationship falls short of statistical significance in this group of flies ( $p = 0.08$ , Pearson's correlation). **b.** There is a trend for the remapping index (in the final part of the inverted gain block) to be negatively correlated with HD encoding accuracy (in the preceding

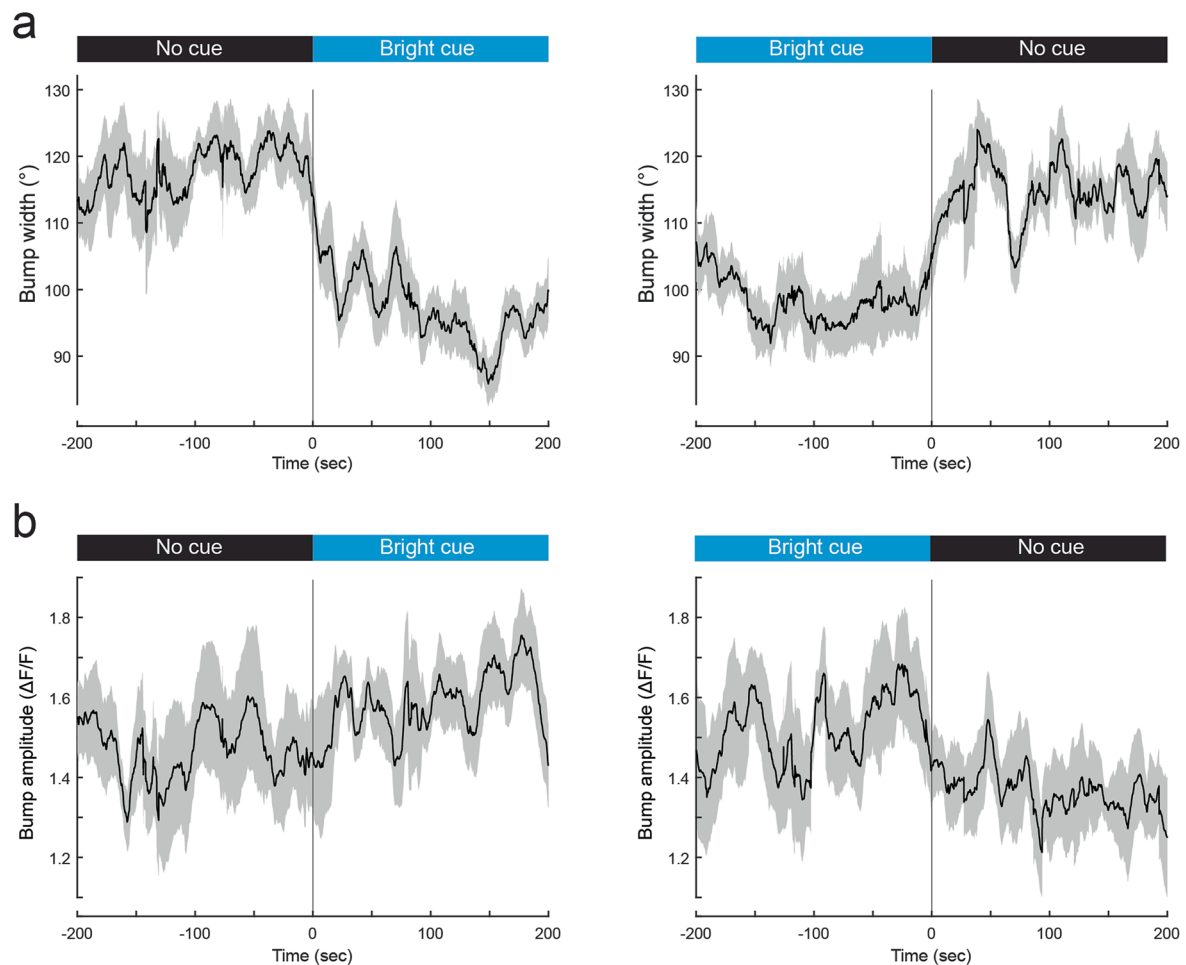
normal gain block), but this relationship falls short of statistical significance ( $p = 0.06$ ). **c.** We finished each experiment by returning to normal gain for 200 s. This panel shows the entire experiment (including this last normal gain block) for one example fly. This is the same example fly shown in Fig. 7b. Note that the EPG bump tracks HD accurately during the final normal gain block. **d.** On average, HD encoding accuracy was no different in the first and last normal gain block ( $p = 0.67$ , paired 2-sided Wilcoxon test).





**Extended Data Fig. 7 | Correlations between bump profile and rotational speed.** Panels **a–d** show data from 28 flies walking in a virtual reality environment with a bright visual cue. Each variable was smoothed with a 10-s moving window, after first excluding time points where the fly was immobile or goodness of fit was below threshold. **a**. Example fly. Here and elsewhere in this figure, we used a 10-s rolling window to compute the rotational speed, bump width and bump amplitude values. **b**. Mean HD encoding accuracy ( $\pm$  s.e.m.) versus the fly's rotational speed. As the fly's rotational speed increases, HD encoding accuracy decreases (linear mixed-effects model,  $p < 0.001$ ). **c**. Mean z-scored bump width ( $\pm$  s.e.m.) versus the fly's rotational speed. As the fly's rotational speed increases, so does bump width (linear mixed-effects model,  $p < 0.001$ ). **d**. Mean z-scored bump amplitude ( $\pm$  s.e.m.) versus the fly's rotational speed. As the fly's rotational speed increases, so does the bump amplitude (linear mixed-effects model,

$p < 0.001$ ). Panels **e–g** show data from 7 flies tested in separate experiments where we displayed a bright visual cue rotating around the fly. The cue rotated at a different speed for each block of the experiment. **e**. Mean cue position encoding accuracy ( $\pm$  s.e.m.) versus cue rotational speed. Cue position encoding accuracy is computed in the same way that we compute HD encoding accuracy, except that it measures encoding of cue position rather than HD. For example, a value of 1 would mean that the EPG bump perfectly tracks the cue, whereas a value of 0 would mean that the bump moves independently from the cue. Cue position encoding accuracy decreases as cue speed increases (linear mixed-effects model,  $p < 0.001$ ). **f**. Mean bump width ( $\pm$  s.e.m.) versus cue rotational speed. Bump width increases as cue speed increases (linear mixed-effects model,  $p = 0.007$ ). **g**. Mean bump amplitude ( $\pm$  s.e.m.) versus cue rotational speed. Bump amplitude decreases as cue speed increases (linear mixed-effects model,  $p < 0.001$ ).



**Extended Data Fig. 8 | Bump profile around the time of a cue change. a.** Bump width over time (mean  $\pm$  s.e.m.), around transitions between darkness and the bright visual cue. This plot shows data from the 15 flies in Fig. 2. Note that bump width decreases abruptly at the time of cue onset, and it increases abruptly at the time of cue offset. **b.** Same but for bump amplitude.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

### Software and code

Policy information about [availability of computer code](#)

Data collection Matlab 2020a, 2021a; ScanImage 2018b or 2020; FicTrac v2.1.1; Python 3.7, 3.9

Data analysis Matlab 2020a, 2021a; Python 3.7, 3.9; RStudio 2022.02.0; R 4.1.3; Stan 2.19.0; PyStan 2.19.0; NoRMCorre

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data are available from the DANDI repository (<https://dandiarchive.org/dandiset/000289>).

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

n/a

Population characteristics

n/a

Recruitment

n/a

Ethics oversight

n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences

☐ Behavioural & social sciences

☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

All sample sizes were chosen based on standard sample sizes in the field. These sample sizes are conventionally determined on the basis of the expected magnitude of animal-to-animal variability, given published results and pilot data. Our sample sizes are similar to those reported in previous publications (refs 23,33,34,37,38,62).

Data exclusions

No flies were excluded from analyses.

Replication

All results reported are based on multiple replicates, with n values stated in the legends. All of the results in the study were found to be reproducible.

Randomization

Following the initial period with a bright cue, the order of the different brightness blocks in the experiment displayed in Figures 2-3 and Extended Data Figures 2 and 8 was randomized, as noted in the Methods.  
The experimental configuration assigned to each fly for the experiment corresponding to figs. 4-5 and sup figs 3-4 (order of wind and visual cues) was also randomized.

Blinding

The experimenter was not blind to the experimental conditions assigned to each fly. Blinding is only potentially relevant to experiments where different flies experienced different experimental treatments (e.g., different genotypes or different drug treatments). However, the was only one experiment in the paper where different flies experienced different experimental treatments, namely the cases where the order of the different brightness blocks was randomized (Figures 2-3 and Extended Data Figures 2 and 8). Here, the experimenter could not be blind to the experimental condition because the stimulus apparatus delivered the stimulus in a manner that was visible to the experimenter.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Involvement in the study  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                             |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology          |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                          |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern           |

### Methods

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Involvement in the study                        |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |



## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	All experimental animals were 1 day old virgin <i>Drosophila melanogaster</i> with the genotype w/+ ; + ; P{R60D05-Gal4}attP2 / P{20XUAS-IVS-jGCaMP7f}VK00005. Both P{R60D05-Gal4}attP2 and P{20XUAS-IVS-jGCaMP7f}VK00005 were obtained from the Bloomington <i>Drosophila</i> Stock Center (RRID:BDSC_39247 and RRID:BDSC_79031, respectively). Virgin female flies were anesthetized on CO2 and collected at least 12 h before the experiment, then allowed to recover on molasses food. Imaging experiments were conducted the following day, 14-36 h after eclosion.
Wild animals	No wild animals were used in this study.
Reporting on sex	Given their larger size, which makes dissections easier, and usually results in better behavior in tethered animals, we relied purely on females for this study. This is a standard practice for researchers in the field conducting neurophysiological experiments in fruit flies.
Field-collected samples	No field samples were collected for this study.
Ethics oversight	No ethical approval was required because experiments were performed on <i>Drosophila melanogaster</i> .

Note that full information on the approval of the study protocol must also be provided in the manuscript.