Irreversible behavior and neural activity in the hippocampus

Kaiyue Shi
Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA;
Department of Physics, Peking University, Beijing 100871, China;
Department of Physics, Yale University, New Haven, Connecticut 06511, USA;
and Quantitative Biology Institute, Yale University, New Haven, Connecticut 06511, USA

Christopher W. Lynn*
Department of Physics, Yale University, New Haven, Connecticut 06511, USA;
Quantitative Biology Institute, Yale University, New Haven, Connecticut 06511, USA;
and Wu Tsai Institute, Yale University, New Haven, Connecticut 06510, USA

(Received 23 January 2026; accepted 27 August 2026; published 15 September 2026)

In the brain, neural activity undergoes directed flows between states, thus breaking time-reversal symmetry. At the same time, animals also exhibit irreversible flows between behavioral states. Yet it remains unclear whether—and how—irreversibility in the brain relates to irreversibility in behavior. Here, we explore this connection in the hippocampus, where neural activity encodes physical location. We show that hippocampal irreversibility can be quantified using the time-delayed cross-correlations between neurons. As a mouse moves along a virtual track, we find that physical flows through the animal's environment generate neural flows through its cognitive map. Strikingly, this neural irreversibility is reproduced by a minimal sensory-representation model with only three parameters: the average velocity of the mouse, the variance in this velocity, and the resolution of the neural encoding. Together, these results provide a mechanistic understanding of irreversibility in the hippocampus and shed light on the links between symmetry breaking in the brain and behavior.

DOI: 10.1103/s4dl-xg31

I. INTRODUCTION

Animals perform sequences of actions that are directed in time, from flocking and foraging to movement, communication, and social interactions [1–10]. Reverse the order of these actions, and the nature of each behavior changes entirely. In this way, animals break time-reversal symmetry by generating net flows between behavioral states [1,11]. Similarly, recent evidence has established that neural dynamics also break time-reversal symmetry across scales. At the level of populations of neurons, collective dynamics exhibit irreversible flows between neural states [12,13]. At the level of coarse-grained neural activity, irreversibility in the human brain is linked to cognition [14,15], consciousness [16,17], and disease [18,19]. Yet despite this progress, we still lack a clear connection between irreversibility in the brain and behavior.

The hippocampus provides an ideal setting to investigate the breaking of time-reversal symmetry [20,21]. A significant fraction of neurons in the hippocampus, known as place cells, each fire when an animal enters a specific location [22,23]. Thus, the collective activity of place cells is thought to represent a cognitive map of the animal's environment, which is critical for spatial processing, episodic memory, and hippocampal replay [24–28]. Given this literal mapping from physical location to neural activity, a clear question emerges: As an animal moves through its environment, do irreversible flows in physical space give rise to irreversible flows in the cognitive map?

II. QUANTIFYING NEURAL FLOWS IN THE HIPPOCAMPUS

To answer this question, we must first quantify irreversibility in the hippocampus. When a place cell fires, this reflects a specific location within the cognitive map, thus defining a neural state [20–23]. To understand the temporal relationships between states, we consider the asymmetric time-delayed cross-correlations,

$$C_{ij}(\tau) - C_{ji}(\tau) = \langle x_i(t)x_j(t+\tau)\rangle_t - \langle x_j(t)x_i(t+\tau)\rangle_t, \quad (1)$$

where $x_i(t)$ represents the activity $[x_i(t) = 1]$ or silence $[x_i(t) = 0]$ of neuron i at time t and $\langle\cdot\rangle_t$ represents an experimental average over time. In the hippocampus, time-delayed correlations are widely applied to characterize neural state changes, such as in hippocampal replay [29] and theta sequences [30–32].

*Contact author: christopher.lynn@yale.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

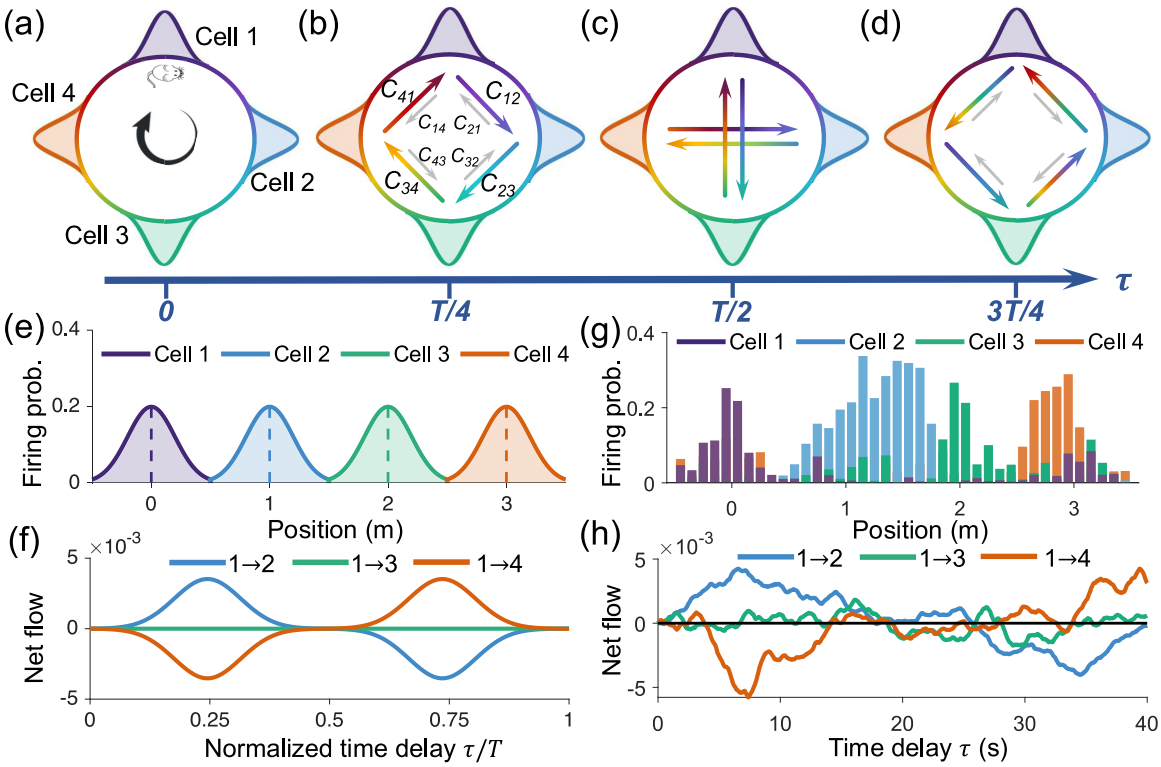

FIG. 1. Flows between neural states in the hippocampus. (a) Place fields for four hypothetical place cells spaced evenly around a circular track as a mouse runs with constant velocity and period T. (b)–(d) Flows between neural states $C_{ij}(\tau)$ for different timescales τ, with gray arrows indicating negligible flows. As the mouse advances along the track, flows progress from clockwise [panel (b)] to directly across the track [panel (c)] to counterclockwise [panel (d)]. (e) Firing probabilities for the hypothetical cells in panels (a)–(d), defined by Gaussians (with uniform variance) spaced evenly along a 4-m track. (f) Net flows $C_{ij}(\tau) - C_{ji}(\tau)$ from cell 1 to cells 2–4 as functions of the normalized time delay τ/T. (g) Place fields for four neurons recorded experimentally from the hippocampus of a mouse as it runs along virtual track of length $L = 4$ m. (h) Net flows $C_{ij}(\tau) - C_{ji}(\tau)$ from cell 1 to cells 2–4 in panel (g) as functions of the time delay τ.

Notably, these correlations can be interpreted as effective flows between neural states. When a place cell fires, this reflects a specific location within the cognitive map, defining a state [20–23]. The time-delayed correlation $C_{ij}(\tau)$ corresponds to the joint probability of being in state i at time t and state j at time $t + \tau$. Thus, the asymmetric $C_{ij}(\tau) - C_{ji}(\tau)$ defines the net flow (or flux) between states on timescale τ. If $C_{ij}(\tau) = C_{ji}(\tau)$ for all neurons i and j—that is, if the correlations are symmetric—then there is no net flow between neural states; reverse the direction of time, and the dynamics remain identical. In contrast, any asymmetry in the correlations defines a net flow $C_{ij}(\tau) - C_{ji}(\tau) \neq 0$ between two locations in the cognitive map, thus breaking time-reversal symmetry.

To illustrate these neural flows, consider a mouse running around a circular track. The irreversibility of the mouse's behavior is self-evident: reverse the sequence of positions, and the mouse runs in the opposite direction. Now, consider four place cells that fire at locations spaced evenly along the track, with firing probabilities represented by the place fields in Fig. 1(a). If the mouse runs with period T, then it moves from one place field to the next on a timescale $\tau = T/4$. This physical movement induces net flows between neural states, breaking time-reversal symmetry in the cognitive map [Fig. 1(b)]. As the timescale increases to a half-period $\tau = T/2$, the flows between different neurons cancel, yielding dynamics that appear entirely reversible [Fig. 1(c)]. Increase the timescale further to $\tau = 3T/4$, and net flows emerge once again, but in the direction opposite the mouse's movement [Fig. 1(d)]. We therefore arrive at the intuition that (1) irreversible movement through an environment may account for irreversible flows in the hippocampus, and (2) these flows should depend critically on the timescale τ [Fig. 1(f)].

III. IRREVERSIBILITY IN THE HIPPOCAMPUS

To test these intuitions, we investigate the activity of 1485 neurons in the CA1 region of the mouse hippocampus, recorded using two-photon calcium imaging (see Appendix A) [33,34]. The mouse runs with its head fixed along a virtual circular track of length $L = 4$ m. Among all neurons, we identify 462 (31%) with statistically significant place fields (see Appendix B), consistent with previous estimates for the density of place cells [20,34]. Consider four such cells with place fields centered approximately equal distances along the track [Fig. 1(g)], thus mirroring our previous example [Fig. 1(e)]. From one cell to the next (in the direction of the mouse's trajectory), we observe an initial peak in the net flow $C_{ij}(\tau) - C_{ji}(\tau)$ followed by a gradual decline to negative flow as τ increases [Fig. 1(h), blue]. Precisely as expected [Fig. 1(f)], this pattern is directly inverted for cells in the reverse direction [Fig. 1(h), orange], and we observe no net flows between cells located on opposite sides of the track

[Fig. 1(h), green]. In addition to the four cells in Figs. 1(g) and 1(h), we observe similar oscillations in the autocorrelations across all place cells in the population (see Fig. S1 in the Supplemental Material [56]).

Rather than studying the flow between each pair of neurons individually, we can quantify the total strength of flows combined across all neurons in the population,

$$\Sigma(\tau) = \frac{1}{2}\sum_{i,j}(C_{ij}(\tau) - C_{ji}(\tau))^2, \tag{2}$$

where the factor of $1/2$ avoids double counting. If the neural dynamics obey time-reversal symmetry on timescale τ, then the net flows between all neurons vanish, and $\Sigma(\tau) = 0$. Meanwhile, any net flow $C_{ij}(\tau) - C_{ji}(\tau) \neq 0$ between a pair of neurons leads to an increase in the total flow $\Sigma(\tau)$. In fact, we show that Σ serves as a lower bound on the information-theoretic irreversibility, which plays a fundamental role in nonequilibrium statistical physics (see Appendix C) [12,35–37]. Finally, note that with no time delay ($\tau = 0$), one will never observe flows in any system, such that $\Sigma(\tau = 0) = 0$ by definition.

Quantifying the total flow between all neurons, we discover that neural dynamics in the hippocampus are highly irreversible [Fig. 2(a)]. Compared to null neurons with activity randomly shifted in time, the real neurons exhibit flows that are up to twice as strong. Moreover, this irreversibility displays clear patterns on three distinct timescales. On a short timescale, the total flow undergoes a sharp increase, reaching a maximum at $\tau = 2.8$ s. On an intermediate timescale, the neural flows appear to oscillate with a period of $\tau \approx 20$ s. Finally, on a longer timescale of $\tau \approx 1$ min, the overall strength of flows decays to the baseline value of the null population. Focusing specifically on the place cells, we find that these patterns are exaggerated [Fig. 2(a)]: The peak in the neural flow becomes sharper, the oscillations are clearer, and the decay is stronger. In contrast, the flows between nonplace cells are significantly weaker and display none of these characteristic features [Fig. 2(b)]. Together, these results demonstrate that irreversibility in the hippocampus arises primarily from the dynamics of place cells. But how are these neural flows related to behavior?

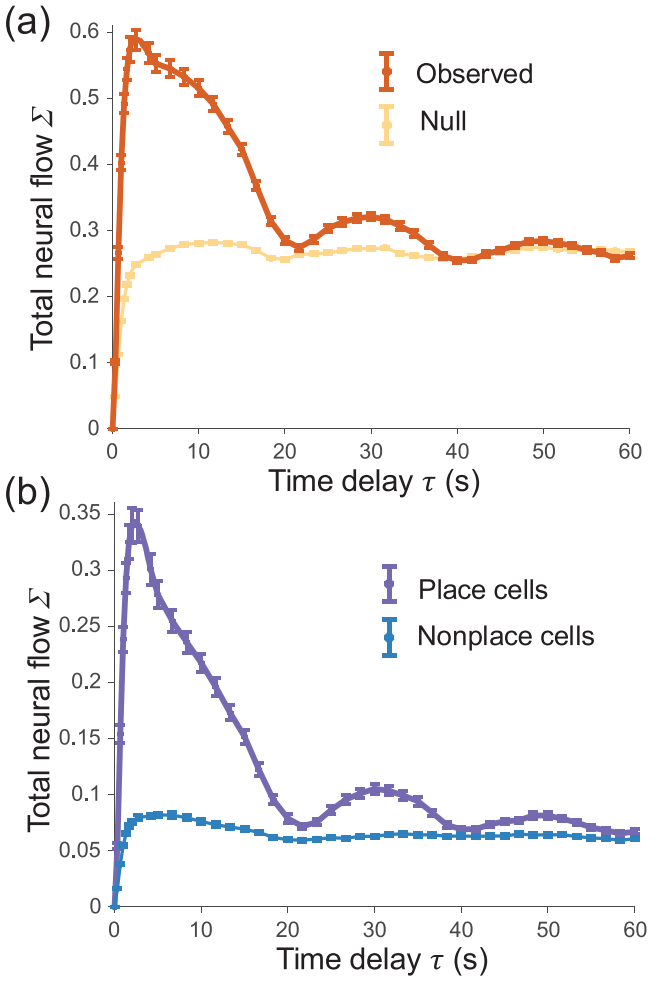

FIG. 2. Neural flow in the hippocampus across timescales. (a) Total neural flow $\Sigma(\tau)$ across a population of 1485 neurons in the hippocampus of a mouse running along a circular track [33,34]. Null values are computed for the same population, but with the activity time series of each neuron shifted by a random length of time (see Appendix A for details). (b) Total neural flow $\Sigma(\tau)$ among the subpopulation of 462 place cells and among the remaining 1023 nonplace cells. Data points and error bars indicate the mean ± 3 SD from 100 bootstrap samples.

IV. VELOCITY ACCOUNTS FOR OSCILLATIONS IN NEURAL FLOW

To understand the relationship between an animal's physical movement and neural flows, we construct a minimal model connecting the two. Consider a population of place cells, each with a place field that defines the firing probability as a function of location. For simplicity, the place field for each cell i is defined by a Gaussian centered at a random position μ_i along the circular track [Fig. 3(a)]. The standard deviation σ, which we hold constant across cells, defines the resolution with which place cells encode the animal's location. If the mouse runs with constant velocity v, then it completes one lap around the track with period $T = \frac{L}{v}$, where L is the track length. For a given pair of cells i and j, we expect the net flow $C_{ij}(\tau) - C_{ji}(\tau)$ to be maximized at the time it takes the mouse to travel from μ_i to μ_j (i.e., $\tau = \frac{\mu_j - \mu_i}{v}$ if $\mu_j \geqslant \mu_i$); conversely, the net flow should be minimized for $\tau = T - \frac{\mu_j - \mu_i}{v}$. Indeed, calculating the net flows (see Appendix E), we observe precisely these maxima and minima, with the pattern repeating for each period T [Fig. 3(b)].

Across all pairs of cells, we find that the net flows vanish at multiples of the half-period $\tau = \frac{T}{2}, T, \frac{3T}{2}, \ldots$ [Fig. 3(b)]. This reversibility in the neural dynamics is directly related to the reversibility of movement in physical space. If we take a snapshot of the mouse's location every half-period, then it appears to complete half a lap at each step, and the trajectory is identical whether viewed forward or backward in time. Given this reversibility in behavior, our model predicts that the total flow $\Sigma(\tau)$ combined across an entire group of place cells should oscillate with a period half that of the animal's physical movement [Figs. 3(c)–3(e)]. Thus, as the mouse runs faster, the total neural flow oscillates with increasing frequency. For a track of length $L = 4$ m and oscillations with period $T \approx 20$ s

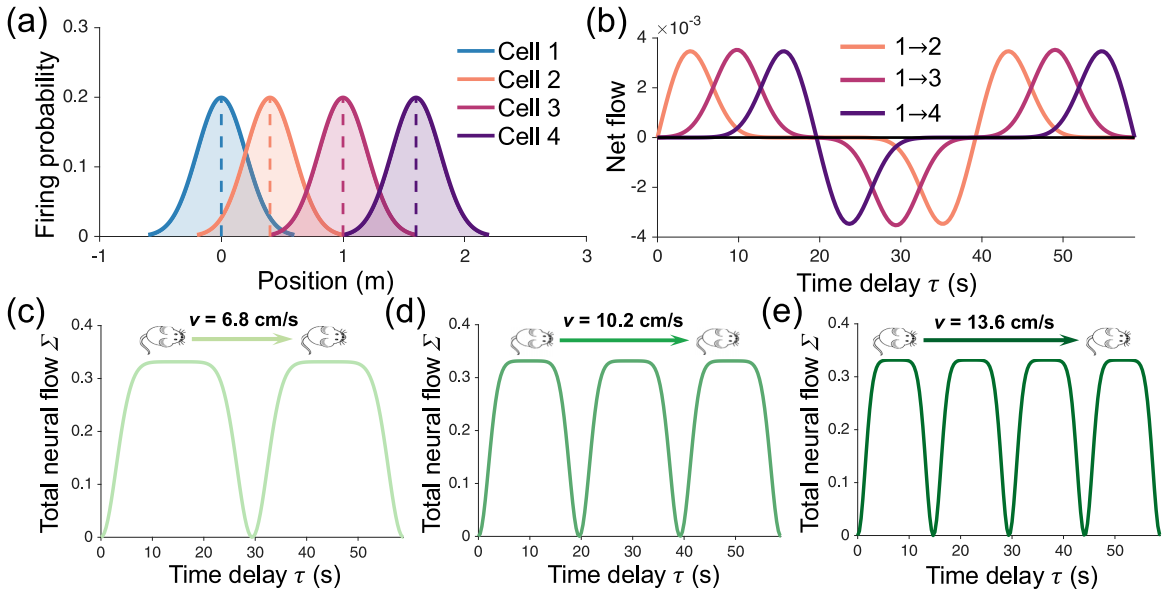

FIG. 3. Mouse velocity determines period of oscillations in neural flow. (a) Example of four place cells in our model, each with a place field defined by a Gaussian centered at a random location along a track of length $L = 4\,\mathrm{m}$ with a common standard deviation $\sigma = 20\,\mathrm{cm}$. (b) Net flows $C_{ij}(\tau) - C_{ji}(\tau)$ from cell 1 to the other three cells for a mouse moving at a constant velocity $v = 10\,\mathrm{cm/s}$, yielding a period $T = 40\,\mathrm{s}$. The net flows between all cells vanish for multiples of $\tau = \frac{T}{2}$, the half-period of the mouse's movement. (c)–(e) Total neural flow $\Sigma(\tau)$ between 462 synthetic place cells for mice running at different velocities v. For $v = 10.2\,\mathrm{cm/s}$ [panel (d)], the period of oscillations matches that measured in experiments (Fig. 2).

(as observed in Fig. 2), our model predicts a velocity of $v = \frac{L}{2T} \approx 10\,\mathrm{cm/s}$ that matches the average speed of the mouse $v = 10.2\,\mathrm{cm/s}$ in the experiment [Fig. 3(d); Appendix G].

V. DIFFUSION ACCOUNTS FOR DECAY IN NEURAL FLOW

We have found that the oscillations in neural flow are consistent with periodicity in behavior, but our model does not yet account for the decay in irreversibility nor the sharp peak (Fig. 2). One might expect the decay in neural flows to arise from noise or heterogeneity in the place cells. To test this hypothesis, we consider Gaussian place fields with random centers μ, random widths σ, random average firing rates r, and additive noise of strength ρ [Fig. 4(a); see Appendix F for details]. Generating a population of such cells [Fig. 4(b)], we simulate activity as the mouse runs at a constant velocity $v = 10.2\,\mathrm{cm/s}$. Surprisingly, the total neural flow does not decay [Fig. 4(c)] and, instead, is nearly identical to the simple model with homogeneous place fields [Fig. 3(d)]. We therefore find that the decay in neural irreversibility is not explained by noise or heterogeneity in the place cells themselves, nor does it arise from correlations in place field locations (see Fig. S2 in the Supplemental Material [56]). Ruling out these potential neural mechanisms, we turn to variability in behavior. Rather than running at a constant velocity, the speed of the mouse naturally fluctuates, which leads to its physical location becoming decorrelated over time. If this physical decorrelation accounts for decorrelation in neural activity, then we should expect the neural flows to decay on long timescales.

To test this hypothesis, we extend our model to include variation in mouse velocity. Specifically, we model the movement of the mouse as a biased random walk with mean velocity v and diffusion coefficient D [38]. Thus, the distance that the mouse travels after an amount of time t is described by a Gaussian $\mathcal{N}(vt, 2Dt)$ that grows wider as t increases [Fig. 4(d)]. With no diffusion ($D = 0$), the mouse runs at a constant velocity, as in our original model (Fig. 3). In this limit, the place cells remain correlated over time [Fig. 4(e)], and we observe no decay in neural flows [Fig. 4(h)]. As diffusion increases to the value $D = 58\,\mathrm{cm}^2/\mathrm{s}$ that best describes the experimental behavior of the mouse (see Appendix G), we find that the place cells become decorrelated over time [Fig. 4(f)]. Combined across an entire population, this leads the total strength of flows $\Sigma(\tau)$ to decrease with increasing time delay τ. As the animal's behavior becomes even more random (increasing D), this decay in neural irreversibility grows stronger [Figs. 4(g) and 4(j)]. Note that despite this decay, the total flow between neurons still oscillates with a period $T/2$, independent from the strength of diffusion D. We therefore find that both the oscillations and decay in neural irreversibility are determined by specific features of the mouse's behavior: average velocity and diffusion.

VI. RESOLUTION ACCOUNTS FOR MAXIMUM NEURAL FLOW

Thus far, our explanations for neural flows have not relied on the properties of the place cells themselves. On short timescales, however, we observe a sharp peak in the irreversibility that is not yet captured by our model (Fig. 2). To understand the timescale τ^* associated with this peak, we must consider the width of the place fields σ, which defines

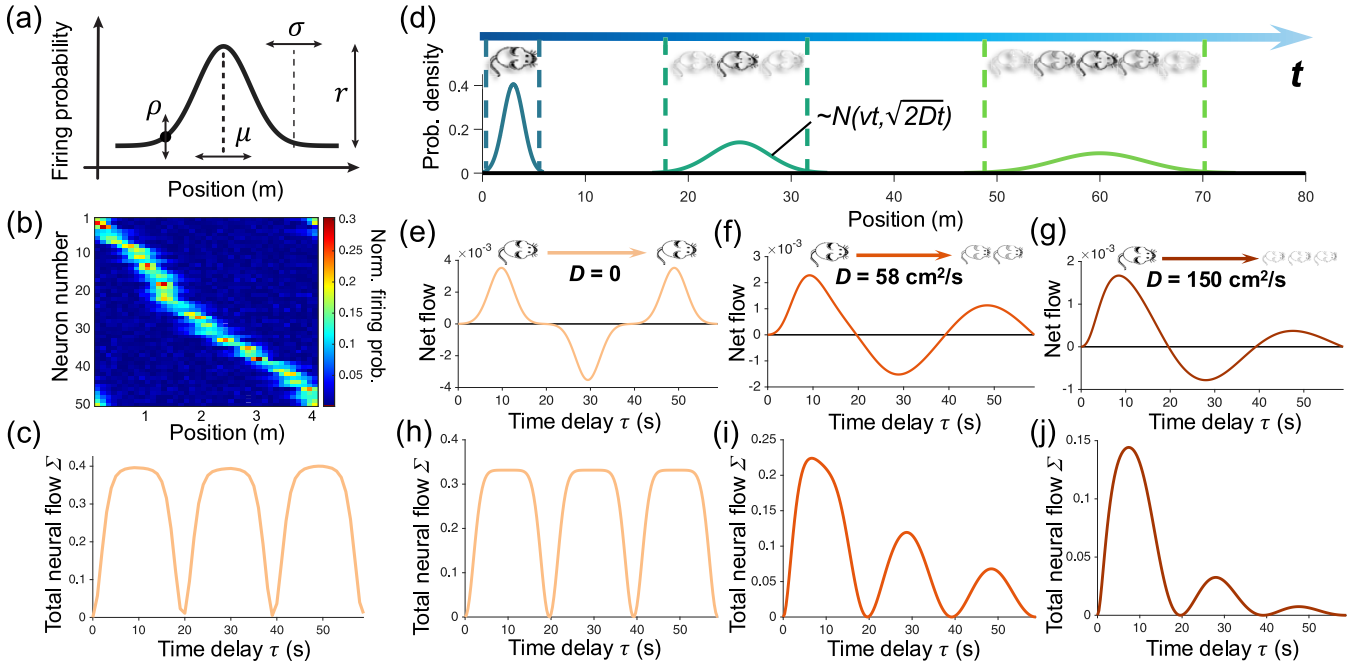

FIG. 4. Diffusion in movement leads to decay in neural flow. (a) We introduce heterogeneity and noise into synthetic place fields via four random parameters: field center μ, field width σ, average firing probability r, and additive noise of strength ρ. (b) Examples of 50 synthetic place fields with randomized parameters defined in Appendix F. Firing probabilities are normalized such that the sum across positions equals one for each cell. (c) Total neural flow for 462 heterogeneous and noisy place cells (matching the number in experiment) simulated as the mouse runs with constant velocity $v = 10.2$ cm/s. (d) Illustration of a biased random walk in which the distance traveled in time t is given by a Gaussian $\mathcal{N}(vt, 2Dt)$, where v is the mean velocity of the mouse and D is the diffusion coefficient. (e)–(g) Net flow between two neurons with place fields centered a distance $\frac{L}{4}$ apart for different diffusion coefficients D. (h)–(j) As D increases, the total flow across an entire population $\Sigma(\tau)$ decays more rapidly. Experimental behavior is best described by $D = 58\,\mathrm{cm}^2/\mathrm{s}$ [panels (f) and (i)]. Across panels (e)–(j), calculations are performed by numerically solving the model with place field width $\sigma = 20$ cm and average velocity $v = 10.2$ cm/s (see Appendix E).

the resolution of the cognitive map. For large σ, it takes longer for the animal to travel from one place field to the next, which is required to produce net flows between neurons [Fig. 5(a)]. In contrast, for small σ the animal will spend less time crossing each place field, and neural flows should increase sharply for short time delays τ.

Decreasing the place field width σ, we find that a pronounced peak emerges in the total neural flow at small τ [Fig. 5(b)]. As the place fields become narrower, increasing their spatial resolution, the maximum flow increases and the associated timescale τ^* decreases [Fig. 5(c)]. Thus, using a minimal model with only three parameters—velocity v, diffusion D, and place field width σ—one can reproduce all of the key features of the neural flows measured in the hippocampus (Fig. 2). Moreover, we confirm that the same neural flows emerge in simulations of mouse behavior and neural activity (see Fig. S3 in the Supplemental Material [56]). Fitting our model to the experimental neural flow curve yields similar parameters to those directly estimated from behaviors and neural data (see Fig. S4 in the Supplemental Material [56]).

For short time delays τ, the model can be solved analytically, yielding an equation relating the timescale of maximum flow τ^* to the three parameters v, D, and σ (see Appendix D). Holding the behavior of the mouse fixed (with v and D set to their experimental values) and sweeping over σ, we confirm that the analytic prediction for τ^* matches the simulation [green line and orange points in Fig. 5(d)]. In the limit of perfect neural resolution ($\sigma = 0$), we find that the peak timescale τ^* does not approach zero; instead, it settles to a minimum value $\tau^*_{\mathrm{min}} = \lambda \frac{D}{v^2}$, where $\lambda = 2.51$ is a constant [dashed line in Fig. 5(d)]. This minimum τ^*_{min} defines the timescale at which the irreversibility of the behavior itself is maximized, without any reference to the cognitive map. Finally, using the analytic relationship between τ^* and σ in our model, we can solve for the place field width that produces the peak timescale $\tau^* = 2.8$ s observed in experiment (Fig. 2). Remarkably, the model predictions are consistent with place fields of width $\sigma \approx 7$ cm, which matches the average width measured in our population of place cells [red point in Fig. 5(d)].

VII. DISCUSSION

Despite growing evidence that the brain breaks time-reversal symmetry across scales, animals, and neural systems [12–19], a concrete connection to behavior has yet to be established. We investigate this connection in the hippocampus, where neural activity defines a cognitive map of an animal's environment [20–23]. Quantifying the flows between neural states, we discover that the hippocampus is highly irreversible. A minimal model directly links this neural irreversibility to animal behavior: As a mouse runs along a circular track, the movement of the animal through physical space generates irreversible flows between neural states in the cognitive map. Using only three parameters—mouse velocity, mouse diffusion, and place cell resolution—this model reproduces the neural irreversibility measured experimentally in the hippocampus.

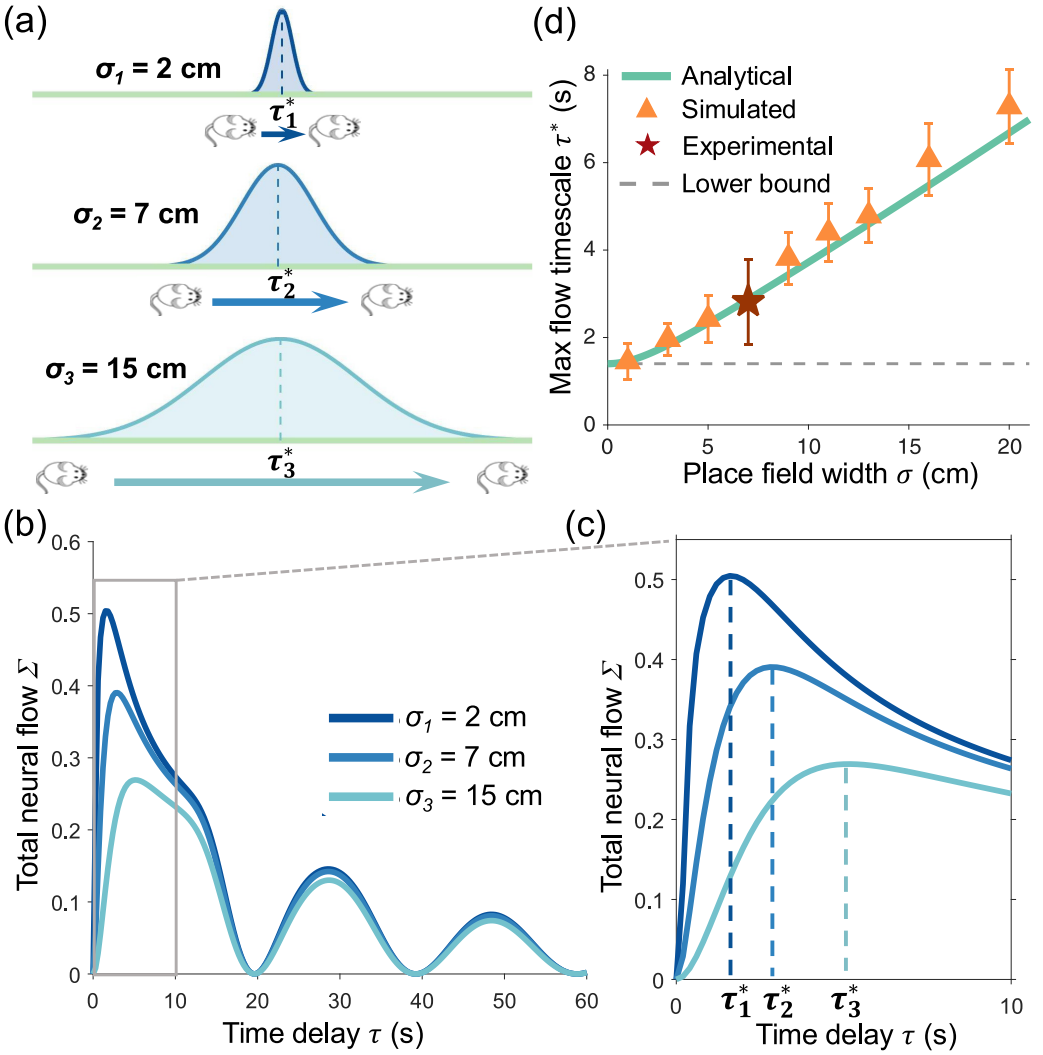

FIG. 5. Resolution of place cells determines peak in neural flow. (a) For wider place fields (larger σ), animals with the same behavior will take longer to travel from one place field to the next, which is required to generate neural flows. (b), (c) Total neural flow $\Sigma(\tau)$ as a function of time delay τ for different place field widths σ. As σ decreases, a peak appears with a higher maximum flow on a shorter timescale τ^*. (d) Timescale of maximum neural flow τ^* as a function of place field width σ. Solid line is the analytic prediction in Eq. (D7). Orange points are simulated τ^* using 1000 synthetic place cells and diffusive trajectories. And dashed line represents the minimum value τ^*_{min} for perfect resolution ($\sigma = 0$). Red point indicates the timescale $\tau^* \approx 2.8$ s and average place field width $\sigma \approx 7$ cm measured in experiment (see Appendix G). Data points and error bars indicate the mean $\pm$ 3 SD from 100 bootstrap samples. Across all panels, the velocity and diffusion of the mouse's movement are set to the experimental values $v = 10.2$ cm/s and $D = 58\,\text{cm}^2/\text{s}$.

While our model predicts the observed neural flows, this does not rule out other sources of irreversibility, such as recurrent neural interactions, internally generated sequences, and hippocampal replay [39,40]. However, the characteristic timescales that we observe in neural flows—particularly the oscillations and decay—are much longer than the timescales intrinsic to neural dynamics [41], and instead precisely match the timescales observed in the mouse behavior. These strong correlations over long timescales suggest that the neural flows are likely externally driven by behavior rather than internally driven by interactions between neurons. Similarly long correlations appear in other neural systems responding to external variables, such as the early visual system responding to repeated stimuli [42]. To directly quantify temporal asymmetries in the dependencies between neurons, future work can extend our analysis to more advanced information-theoretic metrics, such as Granger causality and transfer entropy [43,44].

Given the simplicity of our model, future work can immediately introduce additional biological realism. For example, while we treat place fields as distributed randomly with uniform widths, the firing of place cells can be highly heterogeneous [45,46]. We note, however, that place cell heterogeneity does not explain the observed patterns in the neural flows [Figs. 4(a)–4(c)]. Similarly, while we model mouse movement as a biased random walk, animal behavior is known to exhibit complex correlations across timescales [6,47,48]. The fact that these details are not needed to understand the observed neural irreversibility demonstrates the surprising simplicity of our results. However, investigations into the fine-grained flows between specific neurons will likely require additional biophysical complexity.

More generally, we have only scratched the surface of understanding how external variables drive irreversibility in the brain. As animals move through more complicated environments, are these flows still mirrored in the hippocampus? One can ask similar questions about sequences of abstract concepts and memories, which are also thought to be encoded in the hippocampus [28,49]. In other neural systems, collective activity is tightly linked to motor movements, sensory stimuli, and social processing [50–54]. Does irreversibility in these domains also drive neural irreversibility? Or, alternatively, do neural flows in some systems emerge endogenously from interactions between neurons? Our work provides hope for simple answers to these questions.

ACKNOWLEDGMENTS

We thank J. L. Gauthier for guiding us through the data. We thank M. P. Leighton for helpful discussions about irreversibility. C.W.L. acknowledges support from the NIH (NIH/NIGMS R35GM160188).

DATA AVAILABILITY

The data analyzed in this paper are openly available in Ref. [57].

APPENDIX A: NEURAL DATA

We study patterns of activity in 1485 neurons in the hippocampus of a mouse, recorded in a recent experiment [33]. The mouse is genetically modified so that its neurons express a protein whose fluorescence is modulated by calcium concentration, which in turn follows the electrical activity of the cells. This fluorescence is recorded using a scanning two-photon microscope as the mouse runs along a virtual track of length $L = 4\,\mathrm{m}$. The signal from each cell i consists of a quiet background punctuated by short bursts of activity [34], providing a natural binarization into active $[x_i(t) = 1]$ or silent $[x_i(t) = 0]$ within each video frame t. Capturing images at 30 Hz for 39 min yields $\tau_{\max} = 7.3 \times 10^4$ samples of activity. To generate null data with neural activity that is decoupled from animal position, we shift the time series for each neuron by a random length of time [Fig. 2(a)]. For each neuron, this is equivalent to a random circular permutation in time.

APPENDIX B: IDENTIFYING PLACE CELLS

To identify place cells within the population, we first divide the track into 40 bins of equal width and calculate the probability of activity for each neuron conditional on the animal being located within each bin. For each neuron, we then generate null time series as described above, randomly rotating the activity in time [34]. We repeat this process 1000 times for each neuron and compute the conditional firing probabilities across bins for each permutation. We identify a neuron as a place cell if there are at least three consecutive bins in which the true conditional firing probability is above the 95th percentile across random permutations. This procedure identifies 462 place cells out of 1485 neurons in total (31% of the population), consistent with previous estimates for the density of place cells [20,34].

APPENDIX C: TOTAL FLOW AND IRREVERSIBILITY

Here, we show that the total neural flow in Eq. (2) is a lower bound on the information-theoretic irreversibility, which plays a fundamental role in nonequilibrium statistical physics [12,35–37]. We define the state of the system based on the most recent neuron to fire, such that the number of states matches the number of neurons. The flow from state i to state j on timescale τ is given by the time-delayed cross-correlation $C_{ij}(\tau)$ in Eq. (1). Given a set of flows $C_{ij}(\tau)$, the information-theoretic irreversibility is the Kullback-Leibler divergence between the forward- and reverse-time flows [12–14,36,37]:

$$I(\tau) = \frac{1}{2}\sum_{i,j}(C_{ij}(\tau) - C_{ji}(\tau))\ln\frac{C_{ij}(\tau)}{C_{ji}(\tau)}, \tag{C1}$$

where the factor of $1/2$ avoids double counting. This irreversibility measures the extent to which the dynamics break detailed balance on timescale τ, such that $C_{ij}(\tau) \neq C_{ji}(\tau)$. Note that the flows $C_{ij}(\tau)$ are often normalized to define transition rates or joint transition probabilities; these choices of normalization simply rescale $I(\tau)$ by a constant [12,13,55]. In what follows, we normalize the flows such that $\sum_{ij} C_{ij}(\tau) = 1$; that is, we normalize the flows to define joint transition probabilities.

For a given pair of neurons i and j, let $a = \max\{C_{ij}, C_{ji}\}$ and $b = \min\{C_{ij}, C_{ji}\}$, such that $0 < b \leqslant a < 1$. The corresponding term in Eq. (C1) becomes $(a-b)\ln\frac{a}{b}$. Using the fact that $\ln(1+x) \geqslant \frac{x}{x+1}$ for any x, we have $\ln\frac{a}{b} = \ln(1 + \frac{a-b}{b}) \geqslant \frac{a-b}{a}$. Thus, each term in $I(\tau)$ can be lower bounded, $(a-b)\ln\frac{a}{b} \geqslant \frac{(a-b)^2}{a} > (a-b)^2$. Combining across all pairs of neurons, we have

$$\begin{aligned} I(\tau) &= \frac{1}{2}\sum_{i,j}(C_{ij}(\tau) - C_{ji}(\tau))\ln\frac{C_{ij}(\tau)}{C_{ji}(\tau)} \\ &> \frac{1}{2}\sum_{i,j}(C_{ij}(\tau) - C_{ji}(\tau))^2 = \Sigma(\tau). \end{aligned} \tag{C2}$$

We therefore find that the total flow $\Sigma(\tau)$ in Eq. (2) provides a lower bound on the information-theoretic irreversibility $I(\tau)$.

APPENDIX D: ANALYTICAL MODEL SOLUTION

We model a population of place cells, where the conditional firing probability (i.e., the place field) of each neuron i is proportional to a Gaussian $\mathcal{N}(\mu_i, \sigma^2)$. Explicitly, the firing probability of cell i within each time bin at position s along the track is given by

$$r_i(s) = \frac{1}{k\sqrt{2\pi}\sigma}e^{-\frac{(s-\mu_i)^2}{2\sigma^2}}, \tag{D1}$$

where μ_i is the place field center, σ is the place field width (held constant across all cells), and k is a constant that determines the overall firing probability. Note that the firing probability is defined periodically for positions $\mu_i - \frac{L}{2} + nL \leqslant s \leqslant \mu_i + \frac{L}{2} + nL$, where n is any integer. Adjusting k simply rescales all of our results by a constant factor; we use $k = 10$ for all calculations. We model the movement of the mouse as a biased random walk with average velocity v and

diffusion coefficient D. Thus, the distance d that the mouse travels after time delay τ is also Gaussian distributed,

$$P_\tau(d) = \frac{1}{\sqrt{4\pi D\tau}} e^{-\frac{(d-v\tau)^2}{4D\tau}}. \tag{D2}$$

To compute the time-delayed cross-correlation $C_{ij}(\tau)$, we assume (without loss of generality) that $\mu_i = 0$ and $\mu_j = \delta$ (where $0 \leqslant \delta \leqslant L/2$). Integrating over starting positions $-\frac{L}{2} \leqslant s_i \leqslant \frac{L}{2}$ (at time zero) and all possible final positions s_j (at time τ), we have

$$C_{ij}(\tau) = \int_{-\frac{L}{2}}^{\frac{L}{2}} ds_i \int_{\delta-\frac{L}{2}}^{\delta+\frac{L}{2}} ds_j \times \sum_{n=-\infty}^{\infty} \frac{1}{L} P_\tau(s_j - s_i + nL) r_i(s_i) r_j(s_j), \tag{D3}$$

where the sum in the second equation accounts for the possibility that the mouse travels n laps forward or backward along the track. The physical meaning of each term is clear: The first two terms are the joint probability that the animal is at position s_i at time $t = 0$ and $s_j + nL$ at $t + \tau$; the third term is the firing probability of neuron i at s_i; and the fourth term is the firing probability of neuron j at s_j (which is equivalent to the firing probability at $s_j + nL$).

Each term in the sum in Eq. (D3) defines the contribution to $C_{ij}(\tau)$ from the mouse traveling n laps around the track. On short timescales τ, the mouse is unlikely to travel one lap in either direction, so we can focus on $n = 0$. Additionally, because the place fields are narrow relative to the length of the track, we can approximate the integrals by taking their bounds to infinity, yielding

$$\begin{aligned} C_{ij}(\tau) &\approx \int_{-\infty}^{\infty} ds_i \int_{-\infty}^{\infty} ds_j \frac{1}{L} P_\tau(s_j - s_i) r_i(s_i) r_j(s_j) \\ &= \frac{1}{Lk^2\sqrt{2\pi}\sigma_\tau} e^{-\frac{(v\tau-\delta)^2}{2\sigma_\tau^2}}, \end{aligned} \tag{D4}$$

where $\sigma_\tau = \sqrt{2\sigma^2 + 2D\tau}$. We therefore arrive at an analytic expression for the net flow at short times,

$$C_{ij}(\tau) - C_{ji}(\tau) \approx \frac{1}{k^2 L\sqrt{2\pi}\sigma_\tau} \left(e^{-\frac{(v\tau-\delta)^2}{2\sigma_\tau^2}} - e^{-\frac{(v\tau+\delta)^2}{2\sigma_\tau^2}} \right). \tag{D5}$$

To compute the total neural flow $\Sigma(\tau)$, we average the net flow over the possible differences in place field locations δ and multiply by the number of neuron pairs. Thus, on short timescales, we arrive at the total neural flow,

$$\begin{aligned} \Sigma(\tau) &\approx \frac{N(N-1)}{2} \int_0^{\frac{L}{2}} d\delta \frac{2}{L} (C_{ij}(\tau) - C_{ji}(\tau))^2 \\ &= \frac{A}{\sqrt{\sigma^2 + D\tau}} \left(1 - e^{-\frac{v^2\tau^2}{2(\sigma^2+D\tau)}} \right), \end{aligned} \tag{D6}$$

where $A = \frac{N(N-1)}{2\sqrt{2\pi}k^4L^3}$ is a constant. The maximum total flow is defined by the condition $\frac{\partial\Sigma}{\partial\tau} = 0$, from which we obtain an analytical relationship between the maximum flow timescale τ^* and the parameters in the model,

$$\left(1 + \frac{v^2\tau^*(2\sigma^2 + D\tau^*)}{D(\sigma^2 + D\tau^*)} \right) e^{-\frac{v^2\tau^{*2}}{2(\sigma^2+D\tau^*)}} = 1. \tag{D7}$$

For general v, D, and σ, we cannot solve for τ^* explicitly. However, in the limit of perfect place cell resolution ($\sigma = 0$), we have the minimum timescale

$$\tau^*_{\min} = \lambda \frac{D}{v^2}, \tag{D8}$$

where $\lambda = 2.51$ is a constant [Fig. 5(d)].

APPENDIX E: NUMERICAL MODEL SOLUTION

To compute the correlations $C_{ij}(\tau)$ and total flows $\Sigma(\tau)$ for longer timescales τ, we must include higher-order terms in the sum in Eq. (D3). For each term in the sum, we perform the corresponding integral numerically; we terminate the sum when terms become negligible. Across all analyses, we use $N = 462$ synthetic place cells to match the number identified in experiment.

APPENDIX F: SIMULATED HETEROGENEOUS PLACE CELLS AND NEURAL FLOW

To rule out alternative mechanisms that could explain the decay of neural flow, we consider noisy and heterogeneous place cells. For each synthetic place cell i, the field center μ_i is drawn uniformly at random along the track. Rather than setting all field widths equal, for each cell we draw the width σ_i uniformly at random from $[50\%, 150\%] \times \bar{\sigma}$, where $\bar{\sigma} = 0.2\,\text{m}$ is the field width applied in Figs. 3 and 4. Furthermore, the mean firing probability for each cell r_i is drawn uniformly at random from $[50\%, 150\%] \times \frac{1}{kL}$. Finally, we add noise $\rho_{i,j} = 0.01 \times |\zeta|$ for the jth bin of cell i, where $\zeta \sim \mathcal{N}(0, 1)$. Generating 462 such cells (to match the number in experiments), we simulate activity as the mouse runs at a constant velocity $v = 10.2\,\text{cm/s}$. The resulting neural flows do not decay [Fig. 4(c)].

APPENDIX G: PARAMETER ESTIMATION

Our model only contains three parameters (the average velocity v, the diffusion coefficient D, and the place field width σ), which we can estimate from the experimental data. We estimate the mean velocity v and diffusion coefficient D from the trajectory of the animal's position. First, we remove sections of the trajectory corresponding to the reward zone on the track, where the mouse has an abnormally low velocity (see Fig. S5 in the Supplemental Material [56]). Then, for a given time interval τ, we obtain the distribution of the distances that the mouse has traveled. From this distribution, we compute the mean $\hat{v}(\tau)\tau$ and variance $2\hat{D}(\tau)\tau$ in the distances traveled. We estimate the mean velocity and diffusion coefficient by averaging $\hat{v}(\tau)$ and $\hat{D}(\tau)$ over short timescales from $\tau = 0\,\text{s}$ to $\tau = 4\,\text{s}$. We arrive at the estimated values $\hat{v} = 10.2\,\text{cm/s}$ and $\hat{D} = 58\,\text{cm}^2/\text{s}$. If we include the reward zone in the parameter estimation, then the diffusion coefficient remains nearly identical ($\hat{D} = 59\,\text{cm}^2/\text{s}$), but the estimated velocity decreases ($\hat{v} = 6.5\,\text{cm/s}$), reflecting the fact that the mouse often stops near the reward zone.

We estimate the place field width σ using our place cell detection algorithm. For each place cell, the algorithm detects one or more place fields, each spanning a section of the track. The average length among all detected place fields is

42 cm. In our model, we approximate place fields as Gaussian with standard deviation σ, thus spanning a length of roughly 6σ along the track. We therefore estimate an average place field width of $\sigma \approx 7$ cm. Moreover, by fitting the model's analytical solution to the neural flow curve, we further verify that the parameters closely match those directly estimated from behaviors and neural data (see Fig. S4 in the Supplemental Material [56]).

[1] F. S. Gnesotto, F. Mura, J. Gladrow, and C. P. Broedersz, Broken detailed balance and non-equilibrium dynamics in living systems: A review, Rep. Prog. Phys. **81**, 066601 (2018).

[2] J. Wang, Landscape and flux theory of non-equilibrium dynamical systems with application to biology, Adv. Phys. **64**, 1 (2015).

[3] F. Ferretti, S. Grosse-Holz, C. Holmes, J. L. Shivers, I. Giardina, T. Mora, and A. M. Walczak, Signatures of irreversibility in microscopic models of flocking, Phys. Rev. E **106**, 034608 (2022).

[4] A. Cavagna, I. Giardina, F. Ginelli, T. Mora, D. Piovani, R. Tavarone, and A. M. Walczak, Dynamical maximum entropy approach to flocking, Phys. Rev. E **89**, 042707 (2014).

[5] A. Cavagna, L. Di Carlo, I. Giardina, T. S. Grigera, S. Melillo, L. Parisi, G. Pisegna, and M. Scandolo, Natural swarms in 3.99 dimensions, Nat. Phys. **19**, 1043 (2023).

[6] G. J. Stephens, M. Bueno de Mesquita, W. S. Ryu, and W. Bialek, Emergence of long timescales and stereotyped behaviors in *Caenorhabditis elegans*, Proc. Natl. Acad. Sci. USA **108**, 7286 (2011).

[7] D. W. Stephens and J. R. Krebs, *Foraging Theory* (Princeton University Press, Princeton, 1986), Vol. 6.

[8] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[9] C. W. Lynn, L. Papadopoulos, A. E. Kahn, and D. S. Bassett, Human information processing in complex networks, Nat. Phys. **16**, 965 (2020).

[10] J. Krause, D. Lusseau, and R. James, Animal social networks: An introduction, Behav. Ecol. Sociobiol. **63**, 967 (2009).

[11] E. Schrödinger, *What is Life? The Physical Aspect of the Living Cell and Mind* (Cambridge University Press, Cambridge, 1944).

[12] C. W. Lynn, C. M. Holmes, W. Bialek, and D. J. Schwab, Decomposing the local arrow of time in interacting systems, Phys. Rev. Lett. **129**, 118101 (2022).

[13] C. W. Lynn, C. M. Holmes, W. Bialek, and D. J. Schwab, Emergence of local irreversibility in complex interacting systems, Phys. Rev. E **106**, 034102 (2022).

[14] C. W. Lynn, E. J. Cornblath, L. Papadopoulos, M. A. Bertolero, and D. S. Bassett, Broken detailed balance and entropy production in the human brain, Proc. Natl. Acad. Sci. USA **118**, e2109889118 (2021).

[15] G. Deco, C. W. Lynn, Y. Sanz Perl, and M. L. Kringelbach, Violations of the fluctuation-dissipation theorem reveal distinct nonequilibrium dynamics of brain states, Phys. Rev. E **108**, 064410 (2023).

[16] Y. Sanz Perl, H. Bocaccio, C. Pallavicini, I. Pérez-Ipiña, S. Laureys, H. Laufs, M. Kringelbach, G. Deco, and E. Tagliazucchi, Nonequilibrium brain dynamics as a signature of consciousness, Phys. Rev. E **104**, 014411 (2021).

[17] L. A. de la Fuente, F. Zamberlan, H. Bocaccio, M. Kringelbach, G. Deco, Y. S. Perl, C. Pallavicini, and E. Tagliazucchi, Temporal irreversibility of neural dynamics as a signature of consciousness, Cereb. Cortex **33**, 1856 (2023).

[18] J. Cruzat, R. Herzog, P. Prado, Y. Sanz-Perl, R. Gonzalez-Gomez, S. Moguilner, M. L. Kringelbach, G. Deco, E. Tagliazucchi, and A. Ibañez, Temporal irreversibility of large-scale brain dynamics in Alzheimer's disease, J. Neurosci. **43**, 1643 (2023).

[19] M. Zanin, B. Güntekin, T. Aktürk, L. Hanoğlu, and D. Papo, Time irreversibility of resting-state activity in the healthy brain and pathology, Front. Physiol. **10**, 1619 (2020).

[20] J. O'Keefe and L. Nadel, *The Hippocampus as a Cognitive Map* (Oxford University Press, Oxford, 1978).

[21] J. O'Keefe and J. Dostrovsky, The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat, Brain Res. **34**, 171 (1971).

[22] J. O'Keefe, Place units in the hippocampus of the freely moving rat, Exp. Neurol. **51**, 78 (1976).

[23] E. I. Moser, E. Kropff, and M.-B. Moser, Place cells, grid cells, and the brain's spatial representation system, Annu. Rev. Neurosci. **31**, 69 (2008).

[24] R. G. Morris, P. Garrud, J. a. Rawlins, and J. O'Keefe, Place navigation impaired in rats with hippocampal lesions, Nature (London) **297**, 681 (1982).

[25] J. Bures, A. A. Fenton, Y. Kaminsky, and L. Zinyuk, Place cells and place navigation, Proc. Natl. Acad. Sci. USA **94**, 343 (1997).

[26] D. M. Smith and S. J. Mizumori, Hippocampal place cells, context, and episodic memory, Hippocampus **16**, 716 (2006).

[27] D. J. Foster and M. A. Wilson, Reverse replay of behavioural sequences in hippocampal place cells during the awake state, Nature (London) **440**, 680 (2006).

[28] G. Buzsáki and D. Tingley, Space and time: The hippocampus as a sequence generator, Trends Cognit. Sci. **22**, 853 (2018).

[29] K. Diba and G. Buzsáki, Forward and reverse hippocampal place-cell sequences during ripples, Nat. Neurosci. **10**, 1241 (2007).

[30] W. E. Skaggs, B. L. McNaughton, M. A. Wilson, and C. A. Barnes, Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences, Hippocampus **6**, 149 (1996).

[31] G. Dragoi and G. Buzsáki, Temporal encoding of place sequences by hippocampal cell assemblies, Neuron **50**, 145 (2006).

[32] Y.-H. Yiu, J. K. Leutgeb, and C. Leibold, Directional tuning of phase precession properties in the hippocampus, J. Neurosci. **42**, 2282 (2022).

[33] J. L. Gauthier and D. W. Tank, A dedicated population for reward coding in the hippocampus, Neuron **99**, 179 (2018).

[34] L. Meshulam, J. L. Gauthier, C. D. Brody, D. W. Tank, and W. Bialek, Collective behavior of place and non-place neurons in the hippocampal network, Neuron **96**, 1178 (2017).

[35] U. Seifert, Entropy production along a stochastic trajectory and an integral fluctuation theorem, Phys. Rev. Lett. **95**, 040602 (2005).

[36] É. Roldán and J. M. Parrondo, Estimating dissipation from single stationary trajectories, Phys. Rev. Lett. **105**, 150607 (2010).

[37] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, New York, 2012).

[38] E. A. Codling, M. J. Plank, and S. Benhamou, Random walk models in biology, J. R. Soc. Interface **5**, 813 (2008).

[39] O. L. White, D. D. Lee, and H. Sompolinsky, Short-term memory in orthogonal neural networks, Phys. Rev. Lett. **92**, 148102 (2004).

[40] H. F. Ólafsdóttir, D. Bush, and C. Barry, The role of hippocampal replay in memory and planning, Curr. Biol. **28**, R37 (2018).

[41] J. D. Murray, A. Bernacchia, D. J. Freedman, R. Romo, J. D. Wallis, X. Cai, C. Padoa-Schioppa, T. Pasternak, H. Seo, D. Lee, *et al.*, A hierarchy of intrinsic timescales across primate cortex, Nat. Neurosci. **17**, 1661 (2014).

[42] M. Fritsche, S. G. Solomon, and F. P. de Lange, Brief stimuli cast a persistent long-term trace in visual cortex, J. Neurosci. **42**, 1999 (2022).

[43] A. Shojaie and E. B. Fox, Granger causality: A review and recent advances, Annu. Rev. Stat. Appl. **9**, 289 (2022).

[44] M. Wibral, R. Vicente, and M. Lindner, Transfer entropy in neuroscience, in *Directed Information Measures in Neuroscience*, edited by M. Wibral, R. Vicente, and J. Lizier, (Springer, Berlin, Heidelberg, 2014), pp. 3–36.

[45] R. M. Hayman and K. J. Jeffery, How heterogeneous place cell responding arises from homogeneous grids—A contextual gating hypothesis, Hippocampus **18**, 1301 (2008).

[46] M. I. Anderson and K. J. Jeffery, Heterogeneous modulation of place cell firing by changes in context, J. Neurosci. **23**, 8827 (2003).

[47] W. Bialek and J. W. Shaevitz, Long timescales, individual differences, and scale invariance in animal behavior, Phys. Rev. Lett. **132**, 048401 (2024).

[48] M. P. Leighton and C. W. Lynn, Decomposing non-Markovian history dependence, arXiv:2512.13933.

[49] N. J. Fortin, K. L. Agster, and H. B. Eichenbaum, Critical role of the hippocampus in memory for sequences of events, Nat. Neurosci. **5**, 458 (2002).

[50] K. Svoboda and N. Li, Neural mechanisms of movement planning: Motor cortex and beyond, Curr. Opin. Neurobiol. **49**, 33 (2018).

[51] C. Stringer, M. Michaelos, D. Tsyboulski, S. E. Lindo, and M. Pachitariu, High-precision coding in visual cortex, Cell **184**, 2767 (2021).

[52] M. K. Leonard, L. Gwilliams, K. K. Sellers, J. E. Chung, D. Xu, G. Mischler, N. Mesgarani, M. Welkenhuysen, B. Dutta, and E. F. Chang, Large-scale single-neuron speech sound encoding across the depth of human cortex, Nature (London) **626**, 593 (2024).

[53] K. A. Fulton, D. Zimmerman, A. Samuel, K. Vogt, and S. R. Datta, Common principles for odour coding across vertebrates and invertebrates, Nat. Rev. Neurosci. **25**, 453 (2024).

[54] P. Chen and W. Hong, Neural circuit mechanisms of social behavior, Neuron **98**, 16 (2018).

[55] Q. Yu, M. P. Leighton, and C. W. Lynn, Coarse-graining dynamics to maximize irreversibility, arXiv:2506.01909.

[56] See Supplemental Material at http://link.aps.org/supplemental/10.1103/s4dl-xg31 for extended analysis and discussion.

[57] K. Shi, Neural flow in the hippocampus, GitHub, 2026, https://github.com/Kaiyue-Shi/Neural-Flow-in-the-Hippocampus.