Editors' Suggestion Featured in Physics

Decomposing the Local Arrow of Time in Interacting Systems

Christopher W. Lynn,[1,2] Caroline M. Holmes,[2] William Bialek,[1,2] and David J. Schwab[1]
[1]*Initiative for the Theoretical Sciences, The Graduate Center, City University of New York, New York, New York 10016, USA*
[2]*Joseph Henry Laboratories of Physics and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA*

(Received 8 March 2022; revised 3 June 2022; accepted 24 June 2022; published 6 September 2022)

We show that the evidence for a local arrow of time, which is equivalent to the entropy production in thermodynamic systems, can be decomposed. In a system with many degrees of freedom, there is a term that arises from the irreversible dynamics of the individual variables, and then a series of non-negative terms contributed by correlations among pairs, triplets, and higher-order combinations of variables. We illustrate this decomposition on simple models of noisy logical computations, and then apply it to the analysis of patterns of neural activity in the retina as it responds to complex dynamic visual scenes. We find that neural activity breaks detailed balance even when the visual inputs do not, and that this irreversibility arises primarily from interactions between pairs of neurons.

DOI: 10.1103/PhysRevLett.129.118101

A system held in steady state, away from thermal equilibrium, must continuously dissipate heat to the surrounding bath, causing an increase in entropy. Such a system also violates detailed balance, so that the time reversed trajectories must be measurably less probable than the true trajectories; observation of the system trajectory thus provides evidence for the arrow of time. An important result of modern nonequilibrium statistical mechanics is that the rate at which evidence—in the precise, information-theoretic sense—accumulates for the arrow of time is equal to the rate at which entropy is produced in the bath [1,2]. This idea has been used to search for signatures of irreversibility in experimental data, notably on a wide range of living systems, across scales from single cells [3–5] to global brain dynamics [6,7].

In almost all the systems where we want to study entropy production and the arrow of time, there are many interacting degrees of freedom. If we try to estimate the entropy, then we know that treating each variable independently results in an overestimate, and that as we take account of correlations among pairs, triplets, and larger groups of variables we generate a monotonically decreasing hierarchy of bounds [8]. Here we show that the opposite is true of the entropy production, or the evidence for the arrow of time: we can decompose this measure of irreversibility into a series of non-negative terms, corresponding to successive orders of correlation or interaction among the variables in the system. While correlations always decrease the entropy, we find that correlations always increase the evidence for irreversibility. This leads to a new way of analyzing the origins of irreversibility in interacting systems, which we apply to the neural representation of visual inputs.

Evidence for the arrow of time arises because the probability of observing a trajectory and its time reverse are different, consistently. The proper information-theoretic measure of this difference is the Kullback-Leibler (KL) divergence [9]. If we write trajectories schematically as $x(t)$, with $0 < t < T$, and the corresponding time reversed trajectories as $\tilde{x}(t)$, then the evidence for the arrow of time is

$$E \equiv D_{KL}(P[x(t)]||P[\tilde{x}(t)]) = \sum_{x(t)} P[x(t)] \log\left(\frac{P[x(t)]}{P[\tilde{x}(t)]}\right). \quad (1)$$

For large T in steady-state systems, under mild assumptions, this evidence grows linearly with time, so it is natural to define a (global) rate

$$\dot{I}_{\text{global}} \equiv \lim_{T\to\infty} \frac{E}{T}. \quad (2)$$

For a wide range of systems that can be described as making transitions coupled to a heat bath, this rate is equal to the rate $\dot{S}$ of entropy production that results from heat dissipation to the bath [1,2]. To simplify the discussion, it is convenient to think of time advancing in discrete steps, and to consider observing just one transition, or two steps of the dynamics, at a time. This "local arrow of time" or "local irreversibility" can be written

$$\dot{I} = \sum_{x,x'} P(x \to x') \log\left[\frac{P(x \to x')}{P(x' \to x)}\right], \quad (3)$$

where

$$P(x \to x') = \text{Prob}(x_t = x, x_{t+1} = x'). \quad (4)$$

One can view this not as the Markovian approximation to the global irreversibility, but instead as the *exact* evidence

 © 2022 American Physical Society

for the arrow of time contained in individual transitions. Equation (3) makes very explicit that irreversibility is the breaking of detailed balance.

We are interested in systems where the state x encompasses many interacting variables, $x \equiv \{x_i\}$, with $i = 1, 2, \ldots, N$. If we can measure dynamics with sufficient temporal resolution, then no two variables can change state at the same time. Then instead of considering the full distribution $P(x \to x')$ we can focus on the N individual distributions $P_i(x_i \to x'_i, x_{-i})$, each of which describes a single variable transitioning from x_i to x'_i and the rest of the system remaining in the same state, denoted x_{-i}. Such dynamics are referred to as multipartite, and exhibit a number of useful properties [10–12]. Chief among these properties is the fact that the local irreversibility simplifies to a sum over the irreversibilities associated with the individual elements:

$$\dot{I} = \sum_{i=1}^{N} \dot{I}_i, \tag{5}$$

where

$$\dot{I}_i = \sum_{x_{-i}} \sum_{x_i, x'_i} P_i(x_i \to x'_i, x_{-i}) \log\left[\frac{P_i(x_i \to x'_i, x_{-i})}{P_i(x'_i \to x_i, x_{-i})}\right]. \tag{6}$$

If the different variables in the system are independent of one another, then the dynamics are fully defined by the marginal probabilities $P_i(x_i \to x'_i) = \sum_{x_{-i}} P_i(x_i \to x'_i, x_{-i})$, leading to an irreversibility

$$\dot{I}^{\mathrm{ind}} = \sum_{i=1}^{N} \dot{I}_i^{\mathrm{ind}}, \tag{7}$$

where

$$\dot{I}_i^{\mathrm{ind}} = \sum_{x_i, x'_i} P_i(x_i \to x'_i) \log\left[\frac{P_i(x_i \to x'_i)}{P_i(x'_i \to x_i)}\right]. \tag{8}$$

Even if the variables are not independent we can always define this independent irreversibility. The difference between this and the true irreversibility is the result of interactions,

$$\dot{I}^{\mathrm{int}} \equiv \dot{I} - \dot{I}^{\mathrm{ind}} = \sum_{i=1}^{N} \dot{I}_i^{\mathrm{int}}, \tag{9}$$

where

$$\dot{I}_i^{\mathrm{int}} = \dot{I}_i - \dot{I}_i^{\mathrm{ind}} \tag{10}$$

$$= \sum_{x_{-i}} \sum_{x_i, x'_i} P_i(x_i \to x'_i, x_{-i}) \log\left[\frac{P_i(x_{-i}|x_i \to x'_i)}{P_i(x_{-i}|x'_i \to x_i)}\right] \tag{11}$$

$$= \sum_{x_i, x'_i} P_i(x_i \to x'_i) D_{KL}^i, \tag{12}$$

$$D_{KL}^i = D_{KL}[P_i(x_{-i}|x_i \to x'_i) || P_i(x_{-i}|x'_i \to x_i)], \tag{13}$$

and $P_i(x_{-i}|x_i \to x'_i) = P_i(x_i \to x'_i, x_{-i})/P_i(x_i \to x'_i)$. Since each $D_{KL}^i \geq 0$, this demonstrates that $\dot{I}_i^{\mathrm{int}} \geq 0$, so that interactions can only increase the irreversibility of a system.

Equation (11) admits a simple interpretation: interactions contribute to the arrow of time if the observation of $x_i \to x'_i$ as opposed to $x'_i \to x_i$ points toward different states x_{-i} of the rest of the system. Thus, if i's forward- and reverse-time dynamics contain the same information about the rest of the system, then interactions do not contribute to i's local irreversibility ($\dot{I}_i^{\mathrm{int}} = 0$), and violations of detailed balance can only arise from independent dynamics ($\dot{I}_i = \dot{I}_i^{\mathrm{ind}}$).

Together, Eqs. (7)–(13) establish our first main result: that the local arrow of time can be split into two nonnegative components,

$$\dot{I} = \dot{I}^{\mathrm{ind}} + \dot{I}^{\mathrm{int}}, \tag{14}$$

where $\dot{I}^{\mathrm{ind}}$ reflects the local irreversibility of the individual elements and $\dot{I}^{\mathrm{int}}$ reflects the local irreversibility due to interactions among the elements. Notice that for binary or Ising variables in steady state, we must have $P_i(x_i \to x'_i) = P_i(x'_i \to x_i)$ (such that $\dot{I}^{\mathrm{ind}} = 0$), and so the local arrow of time necessarily arises from interactions ($\dot{I} = \dot{I}^{\mathrm{int}}$). Additionally, we note that decomposition in Eq. (14) requires multipartite dynamics; if multiple elements can change at once, then $\dot{I}^{\mathrm{int}}$ is ill defined (see Ref. [13]).

When we say that interactions contribute to irreversibility, we have the intuition that these contributions can be further decomposed into interactions among pairs, triplets, etc. Saying that we know only about interactions among pairs, for example, is equivalent to saying that we know all the marginal distributions

$$P_i(x_i \to x'_i, x_j) = \sum_{x_{-\{i,j\}}} P_i(x_i \to x'_i, x_{-i}), \tag{15}$$

where the sum runs over the states of all elements other than i and j. We can then ask: what is the minimal irreversibility, or the weakest arrow of time, implied by these pairwise dynamics? To answer this question, we can search over all possible distributions $P_i(x_i \to x'_i, x_{-i})$ that are consistent with the marginals in Eq. (15). Among these hypothetical systems, one will achieve a minimum of the local irreversibility in Eq. (5), thus defining the minimum irreversibility consistent with the observed pairwise dynamics, denoted $\dot{I}^{(2)}$. This generalizes to higher orders, and since knowledge of kth-order dynamics includes all information about dynamics of lower orders $k' < k$, the result is a series of nested bounds,

$$0 \leq \dot{I}^{(1)} \leq \dot{I}^{(2)} \leq \cdots \leq \dot{I}^{(N-1)} \leq \dot{I}^{(N)} = \dot{I}, \qquad (16)$$

where we separately verify that $\dot{I}^{(1)} = \dot{I}^{\mathrm{ind}}$ [13].

The hierarchy of bounds in Eq. (16) is analogous to the hierarchy of bounds on the entropy itself [8], but with inequalities reversed. When subjected to linear constraints on the underlying probability distributions—such as constraints on marginal distributions—entropy has a maximum while mutual information or KL divergence have a minimum [9]. Colloquially, "telling you more" about the distribution increases the evidence for the local arrow of time. Note that computing each bound $\dot{I}^{(k)}$ requires finding a probability distribution that minimizes the irreversibility $\dot{I}$ subject to constraints on the kth-order dynamics [13]; because $\dot{I} = \dot{S}$ in thermodynamic contexts, this is equivalent to asking for dynamics that minimize entropy production. Minimizing entropy production is an idea that has been widely explored [12,14–17], since the foundational work of Onsager and Prigogine [18,19]. Importantly, we are not claiming that real systems minimize their entropy production, but rather are asking for hypothetical systems that have minimal entropy production consistent with a series of increasingly detailed constraints [15,16].

As a final interpretative step, it is natural to compare $\dot{I}^{(k)}$ with $\dot{I}^{(k-1)}$. If $\dot{I}^{(k)} = \dot{I}^{(k-1)}$, then the kth-order dynamics are redundant in the sense that their irreversibility is entirely determined by lower-order correlations; by contrast, if $\dot{I}^{(k)} > \dot{I}^{(k-1)}$, then the kth-order dynamics contain new information about the arrow of time. In this way, we can think about the difference between $\dot{I}^{(k)}$ and $\dot{I}^{(k-1)}$ as the contribution of interactions of order k to the local arrow of time, $\dot{I}^{(k)}_{\mathrm{int}} = \dot{I}^{(k)} - \dot{I}^{(k-1)} \geq 0$. This yields

$$\dot{I} = \underbrace{\dot{I}^{(1)}_{\mathrm{int}}}_{\dot{I}^{\mathrm{ind}}} + \underbrace{\dot{I}^{(2)}_{\mathrm{int}} + \dot{I}^{(3)}_{\mathrm{int}} + \cdots + \dot{I}^{(N)}_{\mathrm{int}}}_{\dot{I}^{\mathrm{int}}}, \qquad (17)$$

which is our central result: the local arrow of time can be decomposed into non-negative contributions from individual elements in the system, interactions between pairs of elements, interactions among triplets, and so on.

To illustrate this decomposition, we consider a minimal system of three binary variables x, y, and z. At each moment in time, z is a noisy logical function of x and y, and from one time step to the next the variables x and y flip between their two states with probability p_{flip}, as in Fig. 1(a). These dynamics are Markovian, so the local arrow of time is also the true global arrow of time. Since the variables are stead state and binary, we have $\dot{I}^{\mathrm{ind}} = 0$, and since there are only three variables, the possible contributions are $\dot{I}^{(2)}_{\mathrm{int}}$ and $\dot{I}^{(3)}_{\mathrm{int}}$.

To begin, consider the simplest function, where z copies either x or y while ignoring the other input [Fig. 1(b)]. As p_{error} increases (that is, as the accuracy of the function decreases), we find that the irreversibility $\dot{I}$ decreases, until at $p_{\mathrm{error}} = 1/2$ (when the output z completely decouples from the inputs x and y) the system becomes reversible ($\dot{I} = 0$). Additionally, the irreversibility vanishes if the inputs x and y are static ($p_{\mathrm{flip}} = 0$) and grows as the inputs become more dynamic (p_{flip} increases). Notably, for all values of p_{flip} and p_{error}, we find that $\dot{I}^{(3)}_{\mathrm{int}} = 0$, and thus the irreversibility of the system arises entirely from pairwise dynamics, $\dot{I} = \dot{I}^{(2)}_{\mathrm{int}}$.

For comparison, consider the AND, OR, and XOR functions [Figs. 1(c) and 1(d)]. As before, the irreversibility increases as the functions become more accurate and as the inputs become more dynamic. In contrast to the copy functions, however, the irreversibilities of AND and OR (which are identical) arise in nearly equal amounts from pairwise and triplet interactions [Fig. 1(c)]. Indeed, for both AND and OR, the output z tends to increase with each of the inputs independently (yielding pairwise irreversibility), yet the full dynamics are not defined until all three variables

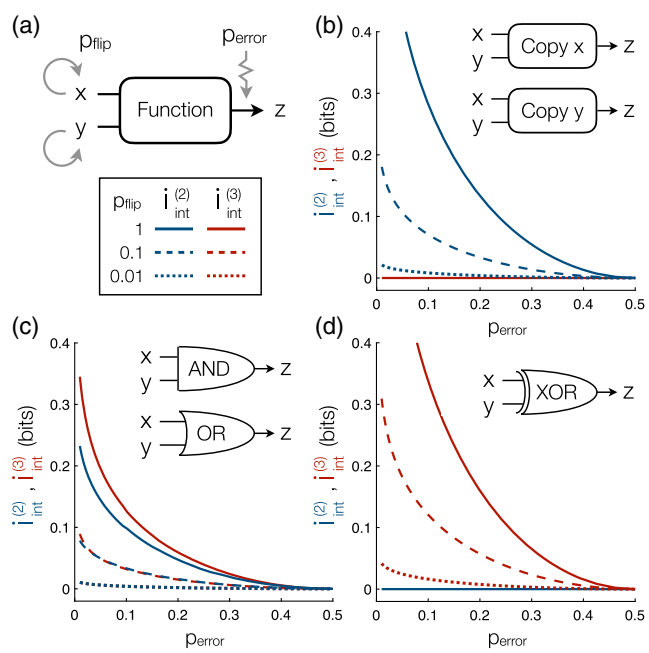

FIG. 1. Decomposing the irreversibility of logical functions. (a) System of three binary variables x, y, and z, where z performs a noisy logical function on the inputs x and y. At each point in time, one of the variables is updated at random. With probability p_{flip}, the inputs x and y change value, and with probability p_{error}, the output z fails to perform the specified function. (b)–(d) Pairwise and triplet contributions to the local arrow of time for different logical functions. Across all functions, irreversibility decreases with p_{error} and increases with p_{flip}. (b) When z copies either x or y, the triplet irreversibility vanishes and all irreversibility arises from pairwise dynamics. (c) For AND and OR, irreversibility is driven by both pairwise and triplet dynamics. (d) For XOR, the pairwise irreversibility vanishes and all irreversibility arises from triplet dynamics.

are taken into account (yielding triplet irreversibility). The XOR function is the classic example of an irreducibly combinatorial interaction, so that the behavior of the system only becomes apparent once all three variables are observed simultaneously. As such, the pairwise dynamics are completely reversible, and all of the irreversibility is driven by triplet dynamics, so that $\dot{I} = \dot{I}^{(3)}_{\rm int}$, as shown in Fig. 1(d). Together, the results in Fig. 1 demonstrate how the local arrow of time can emerge from different orders of interactions among the many variables, even in relatively simple cases.

Our perception of the arrow of time is driven by patterns of activity in networks of neurons. In particular, our experience of the visual world is constructed from the activity of retinal ganglion cells in the optic nerve, carrying information from eye to brain. Experimental developments have made it possible to record, simultaneously, from larger numbers of these cells, to the point that we can monitor all of the information that the brain receives about a small patch of the visual world [20]. Here we focus on one such experiment, monitoring the activity of fifty-three neurons in the salamander retina as it responds to movies with different statistical structure [21]. Neurons have a naturally binary response, generating an action potential (spike; $x_i = 1$) or remaining silent ($x_i = 0$) in each small window of time ($\Delta t = 20$ ms), as in Fig. 2. As we slide the window forward in time, the time resolution of detecting action potentials is high enough that the dynamics are multipartite, so that a spike from only one neuron will enter the front of the window ($x_i = 0 \to 1$) or exit the back ($x_i = 1 \to 0$) at a time, as in Fig. 2(a).

Figure 2(b) shows the probabilities $P(x \to x')$ for a group of three neurons as they respond to a naturalistic movie, and Fig. 2(c) shows the corresponding fluxes of probability between states. Qualitatively we see that there are loops of flux, a hallmark feature of broken detailed balance in steady-state systems [3,4,14,22]. In fact, we have checked that the neurons are in steady state [13], such that the observed irreversibility cannot be attributed to nonstationarity. As with other information-theoretic quantities, estimating the local irreversibility from data is challenging, and prone to systematic errors due to finite data; we find that these can be controlled using the strategy of Ref. [23] if we restrict our attention to groups of no more than five cells. These and other technical details will be addressed in a longer paper [13]. When we choose groups of five cells at random from the fifty-three in the experimental population, more than two-thirds exhibit values of $\dot{I}$ that are significantly different from zero, with the distribution shown in Fig. 2(d).

Movies taken from natural settings, which provide the visual stimuli for the experiments analyzed here, obviously break time-reversal invariance. But one can construct movies for which the time reverse is equally likely, for example, a single bar moving along the trajectory generated

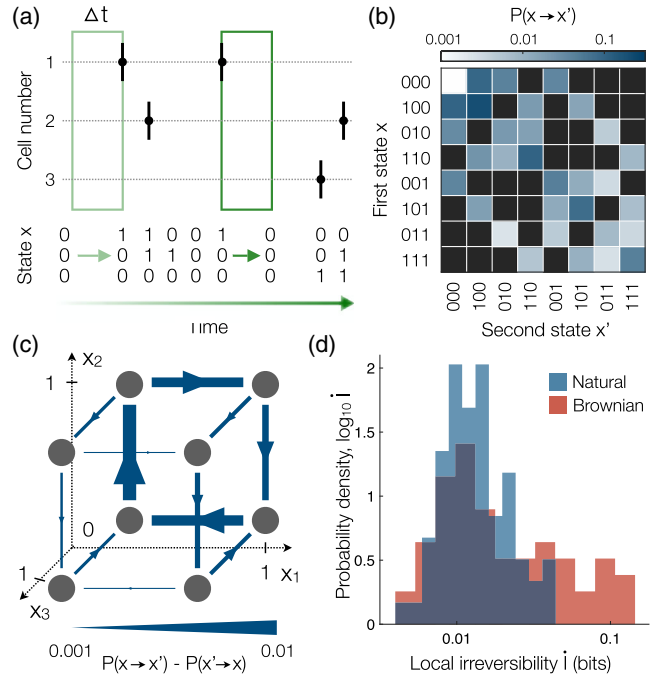

FIG. 2. Local irreversibility in groups of neurons. (a)–(c) A group of three neurons in the salamander retina responding to a natural movie. (a) Points mark the times of action potentials in each of three neurons. To define neural states, we slide a window of width $\Delta t = 20$ ms forward in time. When a cell generates an action potential inside the window we assign $x_i = 1$, and when it is silent we assign $x_i = 0$. Transitions occur when a spike enters the front or exits the back of the window, as shown. (b) The distribution $P(x \to x')$ over pairs of the eight possible states. Black entries indicate disallowed transitions. (c) Fluxes of probability $P(x \to x') - P(x' \to x)$ between neural states, demonstrating that this group of neurons breaks detailed balance. (d) Distributions of local irreversibilities $\dot{I}$ over 5-cell groups responding either to a natural movie (blue) or a Brownian stimulus (red), which is designed to be time-reversal invariant. Out of 100 random 5-cell groups, we only plot those ($\sim 2/3$) with significant local irreversibility.

by equilibrium Brownian motion of a damped harmonic oscillator, as was done (for different reasons) in the experiments of Ref. [21]. Among the same groups of five neurons, we again find that roughly two-thirds exhibit nonzero local irreversibility, but perhaps surprisingly, these groups are more irreversible during the Brownian stimulus than the natural movie [Fig. 2(d)]. This difference in local irreversibility holds for all group sizes from $N = 2$ to $N = 5$. Together, these results demonstrate that retinal neurons break detailed balance in a way that does not simply reflect the irreversibility of the stimulus. In fact, the neuronal dynamics can define an arrow of time, even when the stimulus does not.

The decomposition in Eq. (17) gives us the opportunity to ask how the local arrow of time in neural responses is distributed across dynamics of different order. We recall that the interaction irreversibilities $\dot{I}^{(k)}_{\rm int}$ fundamentally

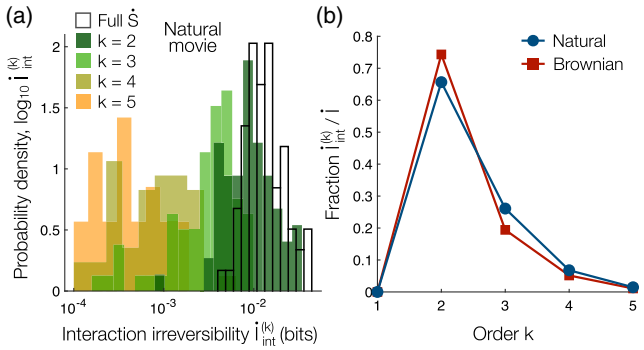

FIG. 3. Decomposing local irreversibility in neuronal activity. (a) Distributions of interaction irreversibilities $\dot{I}_{\rm int}^{(k)}$ of different orders k for random groups of five neurons responding to the natural movie. (b) Interaction irreversibility $\dot{I}_{\rm int}^{(k)}$, normalized by the full local irreversibility $\dot{I}$, as a function of the order k, averaged over the same five-cell groups.

quantify the local arrow of time in kth-order dynamics, and need not be driven by direct connections between neurons themselves. As noted above, the irreversibility that we observe is not the result of nonstationarity; more rigorously we find that in all the data we consider here, distributions really are in steady state within experimental error [13]. But for steady-state binary systems, the individual neurons cannot establish a local arrow of time ($\dot{I}^{\rm ind} = 0$); note that extended sequences of transitions from a single cell could generate irreversibility, but we focus here only on the local term. Thus, any irreversibility necessarily arises from statistical dependencies between two or more neurons at a time. For the same groups of $N = 5$ cells in Fig. 2(d) responding to the natural movie, we find that pairwise dynamics account for much more of the local irreversibility than complex higher-order dynamics [Fig. 3(a)]. In fact, for both the natural and (time-reversal invariant) Brownian movies, pairwise dynamics contribute 66%–74% of the local irreversibility, more than any of the higher order terms [Fig. 3(b)].

To summarize, we have shown how evidence for the local arrow of time accumulates from the behavior of individual degrees of freedom and their interactions. Progressively higher-order dynamics each make a nonnegative contribution, adding to the local irreversibility. As a practical matter, this decomposition allows us to (lower) bound the local irreversibility through measurements of low-order correlations in many-body systems, in much the same way that the maximum entropy method allows us to (upper) bound the entropy itself. We have focused here on the magnitude and decomposition of the local arrow of time, but it would be interesting to explore the hierarchy of minimally irreversible models that we construct along the way, especially as models for living systems. Perhaps these will be as successful in describing dynamics as the maximum entropy models have been in describing distributions of states at single moments in time [24–28]. It would be interesting to understand the relationship of the minimally irreversible models to maximum entropy models for trajectories, sometimes called maximum caliber [29,30].

As a first step we have used our decomposition to analyze the responses of small groups of neurons in the retina as they encode complex visual inputs. It is not surprising that this initial neural representation of the visual world defines an arrow of time, although it is reassuring that this can be quantified, reliably. It is perhaps surprising that irreversibility is stronger in response to inputs that obey detailed balance, raising questions about how our internal perception of the arrow of time becomes aligned with the external world. Despite these large differences in the strength of the local arrow of time in response to different inputs, the way in which large-scale irreversibility is built out of fine-scale dynamics is constant, with the dominant role played by correlations among pairs of neurons. This relative simplicity holds out promise for simplified models of the neural dynamics, similar to pairwise maximum entropy models in the study of steady-state distributions. Generally, the emergence of irreversibility from pairwise dynamics opens the door for future investigations into whether, and how, the physical connections between neurons combine to produce a collective arrow of time.

We thank S. E. Palmer for helpful discussions and for guiding us through the data of Ref. [21]. This work was supported in part by the National Science Foundation, through the Center for the Physics of Biological Function (PHY–1734030) and a Graduate Research Fellowship (C. M. H.); by the National Institutes of Health through the BRAIN initiative (R01EB026943); by the James S. McDonnell Foundation through a Postdoctoral Fellowship Award (C. W. L.); by the Simons Foundation; and by a Sloan Research Fellowship (D. J. S.).

[1] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
[2] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007).
[3] C. Battle, C. Broedersz, N. Fakhri, V. Geyer, J. Howard, C. Schmidt, and F. MacKintosh, Science **352**, 604 (2016).
[4] F. S. Gnesotto, F. Mura, J. Gladrow, and C. P. Broedersz, Rep. Prog. Phys. **81**, 066601 (2018).
[5] T. H. Tan, G. A. Watson, Y.-C. Chao, J. Li, T. R. Gingrich, J. M. Horowitz, and N. Fakhri, arXiv:2107.05701.
[6] C. W. Lynn, E. J. Cornblath, L. Papadopoulos, M. A. Bertolero, and D. S. Bassett, Proc. Natl. Acad. Sci. U.S.A. **118**, e2109889118 (2021).
[7] Y. Sanz Perl, H. Bocaccio, C. Pallavicini, I. Pérez-Ipiña, S. Laureys, H. Laufs, M. Kringelbach, G. Deco, and E. Tagliazucchi, Phys. Rev. E **104**, 014411 (2021).
[8] E. Schneidman, S. Still, M. J. Berry II, and W. Bialek, Phys. Rev. Lett. **91**, 238701 (2003).
[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, New York, 2012).

[10] J. M. Horowitz and M. Esposito, Phys. Rev. X **4**, 031015 (2014).
[11] J. M. Horowitz, J. Stat. Mech. (2015) P03006.
[12] D. H. Wolpert, New J. Phys. **22**, 113013 (2020).
[13] C. W. Lynn, C. M. Holmes, W. Bialek, and D. J. Schwab, companion paper, Phys. Rev. E **106**, 034102 (2022).
[14] J. Schnakenberg, Rev. Mod. Phys. **48**, 571 (1976).
[15] D. Skinner and J. Dunkel, Proc. Natl. Acad. Sci. U.S.A. **118**, e2024300118 (2021).
[16] D. J. Skinner and J. Dunkel, Phys. Rev. Lett. **127**, 198101 (2021).
[17] S. Still, Phys. Rev. Lett. **124**, 050601 (2020).
[18] L. Onsager, Phys. Rev. **37**, 405 (1931).
[19] I. Prigogine, Acad. R. Belg. Bull. Cl. Sci. **31**, 600 (1945).
[20] O. Marre, D. Amodei, N. Deshmukh, K. Sadeghi, F. Soo, T. Holy, and M. J. Berry II, J. Neurosci. **32**, 14859 (2012).
[21] S. E. Palmer, O. Marre, M. J. Berry II, and W. Bialek, Proc. Natl. Acad. Sci. U.S.A. **112**, 6908 (2015).
[22] J. Li, J. M. Horowitz, T. R. Gingrich, and N. Fakhri, Nat. Commun. **10**, 1 (2019).
[23] S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek, Phys. Rev. Lett. **80**, 197 (1998).
[24] E. Schneidman, M. J. Berry II, R. Segev, and W. Bialek, Nature (London) **440**, 1007 (2006).
[25] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, Proc. Natl. Acad. Sci. U.S.A. **109**, 4786 (2012).
[26] L. Meshulam, J. L. Gauthier, C. D. Brody, D. W. Tank, and W. Bialek, Neuron **96**, 1178 (2017).
[27] C. W. Lynn, L. Papadopoulos, D. D. Lee, and D. S. Bassett, Phys. Rev. X **9**, 011022 (2019).
[28] W. P. Russ, M. Figliuzzi, C. Stocker, P. Barrat-Charlaix, M. Socolich, P. Kast, D. Hilvert, R. Monasson, S. Cocco, M. Weigt, and R. Ranganathan, Science **369**, 440 (2020).
[29] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Rev. Mod. Phys. **85**, 1115 (2013).
[30] A. Cavagna, I. Giardina, F. Ginelli, T. Mora, D. Piovani, R. Tavarone, and A. M. Walczak, Phys. Rev. E **89**, 042707 (2014).