

# Decomposing the Local Arrow of Time in the Brain

Researchers have developed a way to quantitatively evaluate irreversibility in complex networks.

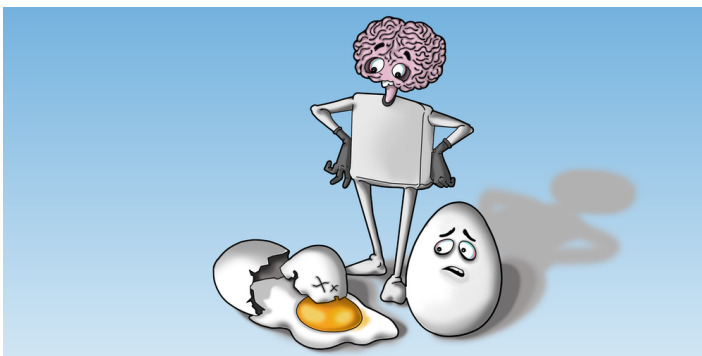
By **Yasser Roudi** and **John Hertz**

A broken egg cannot spontaneously unbreak, and a drop of ink once mixed in water cannot spontaneously unmix. Nature is full of such irreversible phenomena, actions that cannot undo themselves. This irreversibility is quantified by the so-called entropy production rate, which, according to the second law of thermodynamics, is always positive [1]. Thus, one can think of the entropy production rate as a measure of the flow or “arrow” of time for a system. However, measuring this parameter is tricky for complex systems, such as the brain, that have nontrivial, complex interactions between their constituent elements. Now Christopher Lynn of the City University of New York and Princeton University and colleagues present a method to quantify entropy production in such a system [2, 3] (Fig. 1). The team applies its method to the activity of neurons in the retina of a salamander as the system responds to a series of complex visual images. Their work opens the door to the quantitative analysis of the arrow of time in complex biological systems, such as the neuronal networks in the brain, where the

model could potentially enable a quantitative understanding of the neural basis of our perception of the passage of time.

At a basic level, a system’s irreversibility is mathematically equivalent to the “distance” between the probability of the system transitioning between two states and the probability of the reverse transition [4]. Whenever those probabilities are different, the distance is positive. Irreversibility, so defined, is not an all-or-nothing quantity; it can have any positive value. Estimating that value from data from real systems can be very difficult, particularly for biological systems, most of which are highly complex and have many interacting variables. For example, in the brain there are billions of neurons—the brain’s information messengers—which emit tiny voltages to create different voltage-spike patterns—the states of the brain. Even in just a millimeter-sized piece of the brain there are thousands of neurons, making measurements tricky. Current experimental know-how only allows scientists to record with high temporal resolution the spike trains of a few hundred or so neurons. However, the spike trains cannot be recorded for long enough to permit calculating irreversibility for networks of interconnected neurons. There is therefore a need for ways to calculate irreversibility that work for the kind of limited data that we have. This problem is the one that Lynn and his colleagues address.

In their study, Lynn and colleagues show that, in a system such as the brain, irreversibility can be decomposed into a series of terms calculable from first-order statistics, pairwise statistics, triplet statistics, and so on, up to the  $N$ th order, where  $N$  is the number of variables. In the case of the brain, these terms correspond to contributions from spiking statistics of single



**Figure 1:** Researchers have developed a method to calculate the irreversibility of the behaviors of complex systems.

Credit: S. Mohsenin

neurons, pairs of neurons, triplets of neurons, and so on.

The team shows that this decomposition works for calculating the irreversibility of some simple small models, for which they evaluate the entire series. For example, the team applies its model to a system of Boolean logic gates, which perform operations on multiple binary inputs, producing a single binary output. They also consider a theoretical system of neurons that produce spike trains of the length recordable today. In that case, they find that they can accurately estimate the lower-order statistics but not the higher order ones—as the order goes up, more and more data are needed for the calculations and the estimations become impossible. That finding indicates that if the series converges sufficiently rapidly for a given system, it should be possible to reasonably calculate irreversibility using the formalism.

In the case of neural data, Lynn and colleagues also applied their method to spike trains recorded from the retina of a salamander that was subjected to different visual stimuli: a movie of a natural scene and an artificially constructed, reversible movie showing Brownian motion. They show that their data were sufficient to estimate terms up to fifth- or sixth-order statistics, meaning a complete analysis is possible only if a system contains up to five or six neurons. The team then makes the following observations about the system: First, the measured degree of irreversibility depends on the movie shown. The irreversibility is also always positive even if the stimulus is reversible, a finding that indicates that the irreversibility of the spike trains is not simply inherited from the stimulus. Neither of these results is particularly surprising, and the latter is expected given that the retina is a biological machine that responds to but does not exactly copy the statistics of the visual stimuli. Perhaps surprising is the observation that the irreversibility of the spike trains is greater when the stimulus is reversible than when it isn't.

The team also finds that low-order (particularly pairwise) statistics account for most of the total estimated irreversibility.

If this result turns out to hold for a much larger neuron population, it will be good news, as experimentalists could simply use the lower-order statistics to get the information they need without going to higher and higher orders. This possibility is, however, not obvious, as the number of  $N$ th-order correlations increases exponentially with the size of the population, so their effect could be more significant for larger populations [5]. Eventually, it could even be possible to apply the model to networks big enough to have behavioral relevance, allowing researchers to address questions such as whether or how the subjectively perceived passage of time is related to irreversibility in the network dynamics.

Experiments to test these questions are of course not yet possible. Still, with the remarkable technological advancements currently taking place in data collection and in the manipulation of complex systems, that could soon change. The quantitative framework of Lynn and his colleagues will aid in designing and analyzing such experiments.

**Yasser Roudi:** Kavli Institute for Systems Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway

**John Hertz:** Nordic Institute of Theoretical Physics (NORDITA), Stockholm, Sweden

## REFERENCES

1. E. Fermi, **Thermodynamics (Dover Publications, New York, 1936)**.
2. C. W. Lynn *et al.*, “Decomposing the local arrow of time in interacting systems,” **Phys. Rev. Lett.** **129**, 118101 (2022).
3. C. W. Lynn *et al.*, “Emergence of local irreversibility in complex interacting systems,” **Phys. Rev. E** **106**, 034102 (2022).
4. U. Seifert, “Stochastic thermodynamics, fluctuation theorems and molecular machines,” **Rep. Prog. Phys.** **75**, 126001 (2012).
5. Y. Roudi *et al.*, “Pairwise maximum entropy models for studying large biological systems: When they can work and when they can't,” **PLoS Comput. Biol.** **5**, e1000380 (2009).