

**Autobiography Without a Self:
Self-Historical Memory as Holonomy of the
Minimal Realization**

PIATRA . INSTITUTE

July 2026

Abstract

A driven deterministic system (Z, U, F) can be read as carrying a memory of its own history. Fibre the state space over an observable self $x = \pi(z)$, quotient each fibre by future behaviour, and let the input words that return the observable to x act on the quotient as a monoid of closed self-histories. A recent proposal builds nine invariants on this picture and reads them as autobiographical structure, a system’s representation of its past selves. This paper shows the construction is classical. The autobiographical fibre A_x is the output- x fibre of the Nerode minimal realization, and the self-history monoid acting on it is the fibre-restricted transition monoid, which is the Krohn-Rhodes holonomy tile; the nine invariants reduce term by term to automata, semigroup, and Koopman quantities. The proposed differential-geometric invariant, an autobiographical curvature $\Omega = dA + A \wedge A$, is ill-posed: connection holonomy is invertible while the self-history monoid contains the erasures that motivate the whole construction, and the fibres change size from one observable value to the next, so no bundle supports it. The self-versus-environment defect Δ_{ext} measures observability of hidden inputs rather than aboutness; a system that copies its last input scores as maximally self-remembering, and the number depends on the observable one chooses by hand. One question survives the reductions: whether a canonical self can be discovered rather than declared. The autobiographically self-predictable observables are closed under common refinement, proved here, and across 150,964 finite systems they never once fail it; under coarsening they fail generically, in 54,300 of the 65,536 systems with four states and two inputs (82.9%), rising to 99.8% at six states, with a four-state minimal counterexample exhibited in full. The self-predictable observables form a closure system with no canonical coarsest member, a structure that coincides with informational closure. Every numeric claim is reproduced by an accompanying exhaustive computation.

1. A memory that is only bookkeeping

A small chemical or gene-regulatory network, simulated as a system of ordinary differential equations, can be put through a Pavlovian conditioning protocol and made to pass it: pair a neutral input with an effective one, and afterward the neutral input alone drives the response. Fernando et al. (2009) built explicit molecular circuits that learn associatively; Biswas, Manicka, Hoel and Levin (2021) find associative memory across published transcriptional-network models; the bistable multisite-phosphorylation cascade of Markevich, Hornberg and Kholodenko (2004) supplies the hysteresis that stores the acquired state. Levin (2022) reads such results as memory and minimal agency in substrates without neurons. The equations do not change during training. A state variable crosses a separatrix and stays across it, so the same present input meets a different internal configuration and produces a different output. The system’s past is recorded in where it now sits.

From that observation a further construction has been proposed. Treat the system as fibred over its own observable present, $x = \pi(z)$, the “present self”. Over each value x sits the set of hidden states that present as x ; quotient it by future behaviour to get an autobiographical fibre A_x ; let a self-history that departs from x and returns act on A_x , generating a monoid of closed self-histories.

On top of this sit nine invariants, offered as a coordinate-free account of which distinctions among a system's past selves remain causally live, what they represent, and how experiences transform them. The vocabulary is memory, identity, autobiography, self-model.

This paper defends one claim. In the finite deterministic setting the construction actually specifies, the apparatus is a relabelling of theory that already exists. The fibre A_x is the output- x slice of the minimal realization; the self-history monoid is the transition monoid of that realization restricted to the slice, which the Krohn-Rhodes holonomy decomposition already studies; the nine invariants are automata, semigroup, and Koopman quantities under new names. One of the nine is worse placed than the others, being ill-posed rather than merely familiar. The self-versus-world invariant measures a system's failure to observe its own inputs, which is a different thing from what it advertises. What is left standing, after the reductions are made, is a single question the reductions cannot reach: the self-observable π is handed to the construction from outside, and the interesting object would be a self the system determines on its own. Section 6 makes that question precise and settles its lattice-theoretic core by exhaustive computation. The answer is partly negative and locates the one place where something new could still live.

2. The construction

A **driven system** is a triple (Z, U, F) with Z a finite set of states, U a finite input alphabet, and $F : Z \times U \rightarrow Z$ a transition map. Write $\Phi_w(z)$ for the state reached from z by the input word $w \in U^*$, with $\Phi_\epsilon(z) = z$ and $\Phi_{wu}(z) = F(\Phi_w(z), u)$. A **self-observable** is a surjection $\pi : Z \rightarrow X$; its value $x = \pi(z)$ is the present self, and $F_x = \pi^{-1}(x)$ is the fibre over x .

Two hidden states in the same fibre are distinguished only if some future intervention exposes a difference in the observable.

Definition 1 (Autobiographical fibre). *For $z, z' \in F_x$, write $z \sim_x z'$ when $\pi(\Phi_w(z)) = \pi(\Phi_w(z'))$ for every $w \in U^*$. The autobiographical fibre is the quotient $A_x = F_x / \sim_x$.*

A word γ with $\pi(\Phi_\gamma(z)) = x$ for the relevant states returns the observable to its starting value. Such a closed self-history acts on A_x by $T_\gamma([z]) = [\Phi_\gamma(z)]$, and the closed self-histories generate a monoid H_x acting on A_x . The monoid is generally not a group: a self-history can merge two classes and no later history separates them again, so erasure is possible and its action is non-invertible.

The proposal attaches nine invariants to (A_x, H_x) , all taken up to conjugacy of the action: the multiplicity $\mu(x) = |A_x|$; the isomorphism class of $H_x \curvearrowright A_x$; a rank spectrum; a reversibility and consolidation profile; an autobiographical curvature $\Omega = dA + A \wedge A$; a retrieval-depth spectrum; a self-history factorization defect Δ_{ext} ; a historical representation lattice; and a persistence spectrum. The next three sections take these in turn.

3. The fibre is the minimal realization

3.1 Identification of the fibre

The Moore machine (Z, U, F, π) has a future-observation map $\beta_z : U^* \rightarrow X$, $w \mapsto \pi(\Phi_w(z))$. Nerode equivalence identifies z and z' when $\beta_z = \beta_{z'}$; the quotient is the minimal realization M , with an induced output map $\bar{\pi} : M \rightarrow X$ and transition action $\bar{\delta}_w$ (Nerode, 1958; Rutten, 2000). Write $Q_x = \bar{\pi}^{-1}(x)$ for its fibre over x .

Proposition 1. *The relation $\sim_x$ is Nerode equivalence restricted to F_x , and A_x is canonically the fibre Q_x of the minimal realization.*

Proof. For $z, z' \in F_x$ the defining condition of $\sim_x$, equality of $\pi(\Phi_w(z))$ and $\pi(\Phi_w(z'))$ for all w , is equality of β_z and $\beta_{z'}$, which is Nerode equivalence. Because $\beta_z(\varepsilon) = \pi(z)$, Nerode equivalence refines the π -fibre partition, so $\bar{\pi}$ is well defined and no Nerode class straddles two values of π . The quotient map $Z \rightarrow M$ therefore maps F_x onto the Nerode classes with output x , which are the elements of Q_x , and it identifies exactly the $\sim_x$ -equivalent states. Hence $A_x \cong Q_x$. $\square$

This adds nothing. The autobiographical fibre holds no information beyond the minimal realization. It is the deterministic predictive state of the system at the observable x , the state defined by which future observations the past permits, in the sense of the predictive-state representations of Littman, Sutton and Singh (2002) and the causal states of Crutchfield and Young (1989); its minimal predictive formulation is that of Shalizi and Crutchfield (2001).

3.2 The self-history monoid is the holonomy tile

Give $M(\min)$ its transition monoid, the set of maps $\bar{\delta}_w$ for $w \in U^*$ under composition. For the fibre Q_x , collect the restrictions of the fibre-preserving words,

$$M_{Q_x} = \{ \bar{\delta}_w \upharpoonright Q_x : w \in U^*, \bar{\delta}_w(Q_x) \subseteq Q_x \}.$$

Proposition 2. *The self-history action $H_x \curvearrowright A_x$ equals the restricted transition action $M_{Q_x} \curvearrowright Q_x$. The group of units of H_x is the Krohn-Rhodes holonomy group of the tile Q_x .*

Proof. A closed self-history at x is a word w whose transition returns the whole fibre, $\bar{\delta}_w(Q_x) \subseteq Q_x$, which is the condition for T_w to be a total self-map of A_x ; under the identification of Proposition 1, $T_w = \bar{\delta}_w \upharpoonright Q_x$. Fibre-preserving words contain the empty word and are closed under concatenation, and every element of M_{Q_x} is $\bar{\delta}_w \upharpoonright Q_x$ for such a w , so the two monoids of transformations of Q_x coincide. The invertible elements are the permutations of Q_x induced by return words, which are the holonomy group of the tile in the Eilenberg holonomy decomposition of the transition monoid (Krohn and Rhodes, 1965; Eilenberg, 1976). $\square$

The construction slices the minimal automaton by the observable and reads off the return-word monoid on each slice. That the monoid rather than a group is the natural object, because learning

includes forgetting, is a feature of the holonomy decomposition, where the transformation semi-group acting on a tile need not be a group. Forgetting is not incidental.

3.3 The nine invariants

Propositions 1 and 2 push the rest through. Each invariant is computed inside $(M(\min), \bar{\pi}, U)$, with the choice of fibre and, for the last, a baseline input.

invariant	classical object
multiplicity $\mu(x) = A_x $	number of minimal-realization states with output x
monoid action $H_x \curvearrowright A_x$	fibre-restricted transition monoid, the Krohn-Rhodes holonomy tile
rank spectrum	image sizes of monoid elements, graded by word length
reversibility profile	Green's relations, resets, nilpotency index of H_x
autobiographical curvature	ill-posed; residue is non-commutativity of H_x (Section 4)
retrieval depth	shortest distinguishing-word length, the Moore-Hopcroft levels
self/environment defect Δ_{ext}	output-observability with unknown inputs (Section 5)
representation lattice	partition lattice $\Pi_{\mu(x)}$, or the congruence lattice of H_x
persistence spectrum	index and period of a baseline map, its Koopman spectrum

The retrieval-depth spectrum is the ladder of k -equivalence levels in Moore and Hopcroft minimization, the shortest word separating two inequivalent states (Hopcroft, 1971). The representation lattice, read literally as the decodable functions of the past, is the partition lattice $\Pi_{\mu(x)}$ of an $\mu(x)$ -element set, a function of the multiplicity alone; read as the features compatible with the dynamics it is the congruence lattice of H_x , recoverable from the second invariant. The persistence spectrum is the cycle and transient structure of one baseline transformation of a finite set, its Koopman eigenvalues being roots of unity on the cycles and zero off them. Barnett and Crutchfield (2015) supply the input-output version of the underlying machinery, the epsilon-transducer, which reappears in Section 5.

The proposed Canonicity Theorem, that a behaviour-preserving isomorphism takes A_x and H_x to conjugate copies, is the universal property of the minimal realization, the finality of the final coalgebra of the Moore functor (Rutten, 2000). It holds because A_x and H_x are built functorially from behaviour. The three characterization results attached to the proposal are the condition for

trivial fibres ($\mu(x) = 1$ for all x , meaning the observable is already the behavioural state), the condition for a non-trivial tile action, and the non-commutativity of a transition monoid.

4. A curvature that cannot exist

The fifth invariant asks for a connection whose parallel transport around a closed self-history is the holonomy T_γ , and reads its curvature $\Omega = dA + A \wedge A$ as an infinitesimal record of order-dependence. The object is not available.

Proposition 3. *No principal or linear connection has the self-history action as its holonomy, in general. Where a connection exists, its curvature is ordinary curvature and is not determined by the holonomy.*

Proof. Parallel transport of any Ehresmann, principal, or linear connection is invertible, with the transport along the reversed path as inverse (Kobayashi and Nomizu, 1963). A transformation monoid containing a non-invertible element is therefore not a transport family, and H_x contains non-invertible elements whenever some closed self-history merges two classes of A_x , which is the erasure the construction is built to represent. Independently, a fibre bundle over a connected base has homeomorphic fibres, while $\mu(x) = |A_x|$ is allowed to vary with x , so $\{A_x\}$ is not a bundle and admits no connection form. When the fibres are constant and every self-history is invertible, so that H_x is a group, a connection exists, and its curvature is that of the connection, which the holonomy fixes only up to the usual holonomy-curvature gap. $\square$

The well-defined residue is the discrete order-witness: for closed self-histories a, b , whether $T_a T_b = T_b T_a$. Non-commutativity of the self-history monoid records that the order of two episodes changes the resulting state, and it is already recorded by the second invariant. The differential-geometric dressing adds nothing a finite monoid does not already hold.

The instability shows up in the proposal's own showcase. Take the state to run over $\{0, +, -\}$ with a memory register in $\mathbb{Z}_3$, a positive excursion acting on the register by $m \mapsto m + 1 \bmod 3$ and a negative excursion by $m \mapsto 0$. As specified, with the register never read out, the three register values are behaviourally identical, since nothing downstream depends on m ; the fibre over 0 collapses to a single class, $\mu(0) = 1$, and the entire $\mathbb{Z}_3$ vanishes under $\sim_0$. The three past selves survive only once a readout is added, a flag reporting whether $m = 0$. With the readout the object is an ordinary Moore machine whose self-history monoid at 0 is generated by a three-cycle and a constant map, and the Krohn-Rhodes decomposition of that monoid is a two-state reset extended by $\mathbb{Z}_3$. The autobiography is real once a channel conveys it, and it is then a finite automaton.

5. Memory of the World

The seventh invariant promises the sharpest distinction in the proposal, memory a system holds of its own past against memory it holds of its environment. Formalized, it measures something else.

Let the system be driven by inputs from an environment, and fix a self-observable π . The factorization defect $\Delta_{\text{ext}}(\pi)$ vanishes when the observed self-trajectory determines the future self-behaviour: any two input words producing the same π -output trajectory leave the system in $\sim$ -equivalent states.

Proposition 4. $\Delta_{\text{ext}}(\pi) = 0$ if and only if the minimal π -transducer is observable from its output under unknown inputs, equivalently has zero structural crypticity, equivalently synchronizes to its output.

Proof. $\Delta_{\text{ext}}(\pi) = 0$ says the output history is a sufficient statistic for the future-relevant state: the map from output trajectories to $\sim$ -classes is single-valued. That is the statement that the current behavioural state is a function of the past output, which is observability of the transducer from its output when the inputs are not seen, and it is the vanishing of the crypticity of the epsilon-transducer, the residual state information the observed sequence omits (Barnett and Crutchfield, 2015). A machine whose output history pins down its state is one that synchronizes to its output. $\square$

Two consequences remove the invariant's advertised meaning. First, the number is not a measure of aboutness. A machine whose internal register holds a copy of the last input has $\Delta_{\text{ext}} = 0$ whenever that input is visible in the output, so a device that does nothing but log its environment scores as maximally self-remembering. The defect counts hidden inputs; a memory of the self and a memory of the world register alike. Second, the number is not intrinsic. Δ_{ext} is a function of the observable π and the alphabet U , both supplied from outside; the same machine returns a positive defect under one observable and zero under a coarser one. The invariant does not read a property of the system, it reads a property of the cut an analyst has drawn across it. Directed information and transfer entropy (Massey, 1990; Schreiber, 2000) give the stochastic cousins of the same quantity, and they too are conditioned on a chosen partition of the variables into internal and external.

6. Discovering the self

Every reduction so far takes π as given. The construction becomes interesting where π becomes the unknown: a bare driven system with no distinguished observable, and the question of whether it determines a self of its own. A self worth the name should be self-predictable, its own observed history fixing its future observable behaviour, and it should be a genuine coarse-graining rather than the full microstate. Both conditions are available without an external cut.

6.1 Autobiographical closure

Let $\text{Proj}(Z)$ be the partitions of Z ordered by refinement, $\rho \preceq \sigma$ meaning ρ is finer. A partition ρ is a candidate observable; its induced future equivalence $\sim^\rho$ identifies states with equal ρ -output on every input word.

Definition 2 (Autobiographical closure). *A partition ρ is autobiographically closed when, for any two*

input words from a common start with identical ρ -output trajectories, the words leave the system in $\sim^\rho$ -equivalent states. Write $\text{Cl}(Z)$ for the set of closed partitions.

Closure is $\Delta_{\text{ext}}(\rho) = 0$ read as a property of ρ : the observable predicts its own future from its own past. The discrete partition, where the observable is the whole state, is closed, and so is the one-block partition, where the observable is constant. The question is what holds between them, and whether closed observables combine into a canonical one.

6.2 Closure under common refinement

Two observables can be combined two ways. Their common refinement $\rho \wedge \sigma$ observes both, its blocks the nonempty intersections of a ρ -block and a σ -block. Their coarser union $\rho \vee \sigma$ lumps their blocks, its equivalence relation the transitive closure of the union. Observing more preserves closure.

Proposition 5. *If $\rho, \sigma \in \text{Cl}(Z)$ then $\rho \wedge \sigma \in \text{Cl}(Z)$.*

Proof. Write $\tau = \rho \wedge \sigma$, whose output at a state is the pair (ρ -output, σ -output). Then $\sim^\tau = \sim^\rho \cap \sim^\sigma$, since equal τ -output on every future word is equal ρ -output and equal σ -output on every future word. Take two input words from a common start with identical τ -output trajectories. Projecting each trajectory to its first and second coordinates gives identical ρ -output trajectories and identical σ -output trajectories. Closure of ρ makes the endpoints $\sim^\rho$ -equivalent, closure of σ makes them $\sim^\sigma$ -equivalent, so they are $\sim^\tau$ -equivalent. Hence τ is closed. $\square$

The simulation confirms Proposition 5 and finds no exception to it: across the 150,964 systems in the sweep, spanning every system with at most four states and a fixed-seed sample at five and six states, common refinement of two closed observables produced a closed observable every time, with zero violations recorded.

6.3 The coarsening failure

Coarsening is different. Lumping two self-predictable observables generically destroys self-prediction.

The search is exhaustive over all driven systems with at most four states and, for each system, over all partitions, testing every pair of closed partitions for closure of both combinations; five- and six-state systems are sampled under a fixed seed. Under common refinement the closure holds throughout. Under coarsening it fails in most systems, and the failure rate climbs with the number of states.

states	inputs	mode	systems	coarsening fails
2	2	exhaustive	16	0
3	2	exhaustive	729	0
3	3	exhaustive	19{,}683	0

states	inputs	mode	systems	coarsening fails
4	2	exhaustive	65{,}536	54{,}300
5	2	sample	25{,}000	24{,}443
4	3	sample	25{,}000	12{,}244
6	2	sample	15{,}000	14{,}975

At four states and two inputs the closed observables fail to be closed under coarsening in 54,300 of 65,536 systems, 82.9%. The rate rises to 97.8% at five states and 99.8% at six, holding the input alphabet at two. Three states are too few for the failure to appear, since the only partitions between the discrete and the one-block are single merges that leave nothing further to lump into non-closure.

The smallest system that fails has four states and two inputs. State 3 moves to state 1 under input u_0 and stays at 3 under u_1 ; states 0, 1, 2 all move to 0 under either input.

state z	$F(z, u_0)$	$F(z, u_1)$
0	0	0
1	0	0
2	0	0
3	1	3

The observable ρ with blocks $\{0\}, \{1, 2\}, \{3\}$ is closed, and so is σ with blocks $\{0\}, \{1\}, \{2, 3\}$. Their coarser union has blocks $\{0\}, \{1, 2, 3\}$, and it is not closed. Start at state 3. The input u_0 sends it to state 1 and the input u_1 keeps it at state 3; under the union both start and both endpoints lie in the block $\{1, 2, 3\}$, so the two one-letter words produce the same observed trajectory. The endpoints differ in future: a further u_0 sends state 1 to state 0, in the block $\{0\}$, and sends state 3 to state 1, in the block $\{1, 2, 3\}$. Same observed self-history, different future. The self-prediction that ρ and σ each supported is lost when their blocks are merged.

6.4 No canonical self

$\text{Cl}(Z)$ contains the one-block partition and is closed under common refinement, so it is a Moore family, a closure system on the lattice of partitions. It has a closure operator sending any observable to its least closed coarsening, the coarsest resolution one must accept above a given observable to make it self-predictable. That much structure is real and canonical.

The coarsening failure denies the rest. Because two closed observables need not have a closed union, there is no greatest closed observable below a given resolution and no unique coarsest self. The maximal self-predictable coarse-grainings form an antichain, and selecting one among them requires a criterion the closure structure does not provide. A canonical self, discovered from the dynamics alone as the coarsest observable that still predicts itself, does not exist in general.

What does exist, the closure operator on self-predictable observables, is the deterministic form of a known construction. Informational closure, the property that a coarse-graining's future is conditionally independent of the fine detail given its own past, generates the same Moore-family structure and the same least-closed-refinement operator (Bertschinger, Olbrich, Ay and Jost, 2008; Pfante, Bertschinger, Olbrich, Ay and Jost, 2014). The tractable half of the discover-the-self problem is informational closure in a deterministic costume. The half that would have been new, a canonical coarse self, is the half the computation rules out.

7. What would be new

The reductions leave one question standing, and it is narrow. A discovered self, however it is selected from the antichain of maximal self-predictable coarse-grainings, could be compared against the partitions that other theories draw for the same purpose: the internal states of a Markov blanket (Friston, 2013; Kirchhoff, Parr, Palacios, Friston and Kiebel, 2018), the components an individuality measure assigns to a system rather than its environment (Krakauer, Bertschinger, Olbrich, Flack and Ay, 2020), the informationally closed coarse-graining. If these always agree, the self-history holonomy adds nothing on top of them. If some finite system exists where a discovered self and a Markov blanket disagree about which states are internal memory, and where the self-history monoid on the discovered self's fibres separates systems the blanket cannot, that disagreement is the first content the construction would own. The structural coincidence of Section 6, the closure operator being informational closure, makes agreement the likely case and lowers the odds that the disagreement exists. The critiques of blanket-based individuation (Biehl, Pollock and Kanai, 2021) are the right adversary for that search, since they already exhibit systems where the blanket partition is non-unique or misleading.

The account here is deterministic and finite. The stochastic reduction of Δ_{ext} to conditioned directed information is stated but not proved at the generality of Section 5's deterministic result, and the identification of the discovered-self closure operator with informational closure is argued structurally rather than by an isomorphism of constructions. Both are the reductions to complete before the disagreement search is worth running.

The result is a deflation with a fixed point. The autobiographical reading of a driven system, the fibre of past selves, the holonomy of remembered episodes, the curvature of experience, resolves into the minimal realization, its transition monoid, and the Krohn-Rhodes holonomy of a tile, with one invariant ill-posed and one measuring the wrong thing. The construction is a lens, and a usable one: fibring a system over its own output and reading return-word holonomy as memory is a legitimate way to organize what a finite machine records of its history. It is a way of seeing classical objects. The self, entered as a primitive through the observable π , turns out to be either handed in from outside or, when the system must find it alone, without a single canonical form. A system remembers its history. It does not, on this evidence, thereby have a self.

References

- Barnett, N., and Crutchfield, J. P. (2015). Computational mechanics of input-output processes: Structured transformations and the epsilon-transducer. *Journal of Statistical Physics*, 161(2), 404–451.
- Bertschinger, N., Olbrich, E., Ay, N., and Jost, J. (2008). Autonomy: An information theoretic perspective. *Biosystems*, 91(2), 331–345.
- Biehl, M., Pollock, F. A., and Kanai, R. (2021). A technical critique of some parts of the free energy principle. *Entropy*, 23(3), 293.
- Biswas, S., Manicka, S., Hoel, E., and Levin, M. (2021). Gene regulatory networks exhibit several kinds of memory: Quantification of memory in biological and random transcriptional networks. *iScience*, 24(3), 102131.
- Crutchfield, J. P., and Young, K. (1989). Inferring statistical complexity. *Physical Review Letters*, 63(2), 105–108.
- Eilenberg, S. (1976). *Automata, Languages, and Machines, Volume B*. New York: Academic Press.
- Fernando, C. T., Liekens, A. M. L., Bingle, L. E. H., Beck, C., Lenser, T., Stekel, D. J., and Rowe, J. E. (2009). Molecular circuits for associative learning in single-celled organisms. *Journal of the Royal Society Interface*, 6(34), 463–469.
- Friston, K. (2013). Life as we know it. *Journal of the Royal Society Interface*, 10(86), 20130475.
- Hopcroft, J. (1971). An $n \log n$ algorithm for minimizing states in a finite automaton. In *Theory of Machines and Computations* (pp. 189–196). New York: Academic Press.
- Kirchhoff, M., Parr, T., Palacios, E., Friston, K., and Kiebel, S. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface*, 15(138), 20170792.
- Kobayashi, S., and Nomizu, K. (1963). *Foundations of Differential Geometry, Volume I*. New York: Interscience.
- Krakauer, D., Bertschinger, N., Olbrich, E., Flack, J. C., and Ay, N. (2020). The information theory of individuality. *Theory in Biosciences*, 139(2), 209–223.
- Krohn, K., and Rhodes, J. (1965). Algebraic theory of machines. I. Prime decomposition theorem for finite semigroups and machines. *Transactions of the American Mathematical Society*, 116, 450–464.
- Levin, M. (2022). Technological approach to mind everywhere: an experimentally-grounded framework for understanding diverse bodies and minds. *Frontiers in Systems Neuroscience*, 16, 768201.
- Littman, M. L., Sutton, R. S., and Singh, S. (2002). Predictive representations of state. In *Advances in Neural Information Processing Systems 14* (pp. 1555–1561). Cambridge, MA: MIT Press.

- Markevich, N. I., Hornberg, J. B., and Kholodenko, B. N. (2004). Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *Journal of Cell Biology*, 164(3), 353–359.
- Massey, J. L. (1990). Causality, feedback and directed information. In *Proceedings of the International Symposium on Information Theory and its Applications (ISITA-90)* (pp. 303–305).
- Nerode, A. (1958). Linear automaton transformations. *Proceedings of the American Mathematical Society*, 9(4), 541–544.
- Pfante, O., Bertschinger, N., Olbrich, E., Ay, N., and Jost, J. (2014). Comparison between different methods of level identification. *Advances in Complex Systems*, 17(2), 1450007.
- Rutten, J. J. M. M. (2000). Universal coalgebra: a theory of systems. *Theoretical Computer Science*, 249(1), 3–80.
- Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85(2), 461–464.
- Shalizi, C. R., and Crutchfield, J. P. (2001). Computational mechanics: pattern and prediction, structure and simplicity. *Journal of Statistical Physics*, 104(3), 817–879.