

CLOSURE UNDER PROJECTION

What a Point of View Can Determine

Flyxion

Independent Researcher

October 2026

Abstract

Miranker and Zuckerman's *Dynamics of mental activity* represents recurrent neural activity by circular set structures and transports threshold-network dynamics into that representation. Under Aczel's anti-foundation axiom each node of a weighted network receives a set, its decoration, built from the decorations of the nodes it connects to together with a well-founded label recording the active incoming weights. The authors show that, for finite graphs in which every node starts an arrow, the label can be read back from the decoration, and that the voltage, histogram and decoration dynamics can then be carried into one another at fixed weights. They go on to interpret the decoration of a single distinguished node, the thema, as a primary experience on which a consciousness operator acts.

This essay separates three things the construction can be taken to show: that a representation is lossless, that a projection of it is dynamically closed, and that the projected object is an experience. The first is established under stated structural restrictions, and what it recovers is the labeling on a fixed graph, not the graph. The second is a checkable condition, $\pi S = F\pi$, which asks whether states that look the same from the projection also look the same after the system updates. Whether it holds must be established for the network and distinguished node under consideration, and the passage from a full decoration to the thema is many-to-one, so the condition can fail where the representation is lossless. The third requires an empirical bridge that the axioms do not supply, since the four axioms for consciousness operators fix a family of selection rules and nothing that separates conscious from unconscious activity.

Treating a point of view as a projection turns the question of subjectivity into the question of when the viewpoint determines its own continuation. Where closure holds, the thematic level is a dynamical system in its own right and can be studied in its own terms. Where it fails, any description confined to the viewpoint depends on information the viewpoint does not contain. The same condition applies to any mechanism that selects which features of a system to retain, and it separates selection that is shown to be sufficient for a task from selection that merely happens to be.

1. CIRCULAR STRUCTURE AND DECORATION

The construction begins with a point about what set membership can express. In ordinary well-founded set theory, membership cannot circle back on itself: no set contains itself, either directly or through a chain of memberships. Aczel’s anti-foundation axiom [1] removes this restriction and in exchange guarantees that every directed graph has a unique decoration, an assignment of a set to each node such that the elements of the set at a node are exactly the sets assigned to its children. The simplest case is the two-node graph with edges $x \rightarrow y$ and $y \rightarrow x$, whose decoration satisfies $d(x) = \{d(y)\}$ and $d(y) = \{d(x)\}$ and so describes mutual membership. Under anti-foundation the unique decoration identifies the two sets, so that $d(x) = d(y) = q$ with $q = \{q\}$, a set known as a Quine atom. Two nodes that are plainly distinct as nodes of the graph denote the same set. A representation by membership structure therefore preserves exactly the distinctions that membership structure can express and discards the others. Labels can break the identification: in the labeled equation introduced below, two such nodes receive the same decoration only if their labels agree.

Miranker and Zuckerman [2, 3] apply this apparatus to recurrent networks by assigning to each node a a set $d(a)$, its decoration, satisfying what they call the M-Z equation

$$d(a) = \{d(b) \mid a \rightarrow b\} \cup H(a),$$

where $a \rightarrow b$ denotes a directed edge and $H(a)$ is a well-founded label encoding a histogram of the active incoming synaptic weights. The first part of the right-hand side collects the decorations of the nodes that a connects to, and the second carries local numerical information about activity. Because the network may contain cycles, these equations cannot be evaluated by working backward from terminal nodes. They must be solved together, which is what the anti-foundation axiom licenses. A pointed graph decorated in this way pictures a set. The authors call such a picture a meme and the set decorating its point the thema. The terms are technical and unrelated to the popular sense of “meme.” A thema can be pictured by many different memes: the foundations paper notes that an unbounded number of neural states and associated graphs can share one and the same thema.

What this representation gains is a description of each node through its relations to the rest of the network, with activity data built in. What it raises is the question that organizes the rest of this essay: which distinctions survive a given representation, and do the surviving distinctions suffice to determine what happens next. The argument proceeds from the representation itself to what can be recovered from it (Section 2), to the dynamics transported through it (Section 3). It then turns to the closure condition for a projection (Section 4), before asking what that condition does and does not license for the interpretation of the thema as experience (Sections 5 and 6).

2. RECOVERABILITY

The dynamics paper's most useful result for present purposes concerns recoverability [3]. It introduces the selector

$$W(x) = \{ y \in x \mid y \text{ is well-founded} \},$$

which keeps only the well-founded elements of a set. Under the restrictions the authors impose in the dynamics paper, every node decoration is non-well-founded, while the elements of the histogram labels are well-founded, the latter because a histogram is a finite set of pairs consisting of a weight and a positive multiplicity [2]. The two components of the decoration equation can then be separated, because W discards the decorations contributed by neighboring nodes and retains the local label. The result is

$$W(d(a)) = H(a),$$

so the histogram labels can be read back out of the encoding. The labels are not the activity data itself, since a histogram already discards which presynaptic neurons were active.

The restriction carries real weight and should be stated as part of the result. If a neighboring decoration were well-founded, W would retain it alongside the label and the simple separation would fail. The recovery theorem is a property of a particular class of decorated graphs, and the authors restrict their development to that class in order to obtain it.

It is also worth being exact about what recovery establishes. That a label can be extracted from a decoration is a fact about the encoding. It does not show that any system consulted the label in producing a decision, used it appropriately, or retained every distinction a later revision would need. Encoding, access, use, and sufficiency are four separate questions, and the paper's recovery result bears only on the first.

Proposition 2.1 (Label recovery under separation). *Let $C_a = \{ d(b) \mid a \rightarrow b \}$ and suppose that every element of C_a is non-well-founded and every element of $H(a)$ is well-founded. Then $W(d(a)) = H(a)$.*

Proof. By the decoration equation, $d(a) = C_a \cup H(a)$. Selecting the well-founded elements discards every element of C_a and keeps every element of $H(a)$, so $W(d(a)) = H(a)$. $\square$

The dynamics paper does not assume the separation directly but obtains it from structural hypotheses. Its Definition 6.1 calls a multigraph non-well-founded when every node is the starting point of an arrow, and Hypothesis 9.2 restricts attention to finite graphs of this kind. Under that restriction every node decoration is non-well-founded, and for any labeling by well-founded sets the labeling is recovered as the

well-founded part of the decoration [3, Propositions 6.2, 6.3 and 9.3]. The proposition above isolates the separation step inside that argument. The restriction is not idle: the paper’s own example of a node with no outgoing arrow has a well-founded decoration, and there the well-founded part picks up that decoration along with the label. What is recovered is the labeling on a fixed graph, not the graph itself.

3. TRANSPORTED DYNAMICS

The dynamics begin from the McCulloch–Pitts model of a binary threshold neuron [2, 3]. The voltage of a neuron at the next step is

$$v_{\text{new}}(a) = h\left(\sum_{p: p \rightarrow a} w(p \rightarrow a) v_{\text{old}}(p) - \theta\right),$$

where h is the Heaviside function, θ is a threshold, and the sum runs over the neurons p that connect directly to a . Incoming weighted activity decides whether a neuron fires. The foundations paper pairs this with Hebb’s rule for weight change, so that the complete model is a coupled system for voltages and weights; the dynamics paper itself takes the synaptic weights to be time independent, so that the graph, the weights and the threshold are parameters of every update it defines, and the same convention is followed here [3, Section 7 and Remark 7.5]. The transported dynamics therefore describe changing activity at fixed connectivity, and not the coupled evolution of activity and plasticity. From the voltage update the dynamics paper builds dynamics for the histograms, generalized to what it calls crops, and then for the decorations.

Write A_Γ for the transformation from labels to decorations and T for the update on labels induced by the voltage update at fixed weights. The induced update on decorations is

$$S = A_\Gamma T A_\Gamma^{-1}.$$

On the relevant image, W supplies the inverse, which gives

$$d(t+1) = A_\Gamma T W d(t).$$

This is transport of dynamics through an invertible representation: recover the labels, update them, and rebuild the decoration. It shows that the set-based description evolves consistently with the underlying activity model, and nothing more. Nor does the invertibility of the representation make the evolution reversible in time. The threshold update can send distinct activity patterns to the same pattern, and a lossless change of representation inherits that. Invertibility of the encoding and reversibility of the process are different properties, and the paper establishes only the first.

4. THE CLOSURE CONDITION

Closure here means dynamical closure of a projection: whether the projected states can be updated without reference to the full state they were drawn from.

Definition 4.1 (Dynamical closure of a projection). Let S update full states and let π extract a projected state. A deterministic update F on projected states exists with

$$\pi S = F\pi$$

if and only if

$$\pi(\mu_1) = \pi(\mu_2) \implies \pi(S\mu_1) = \pi(S\mu_2).$$

The projection π is then *closed* under S .

This is the congruence condition behind lumpability of Markov chains [4], bisimulation quotients [5], and homomorphisms of dynamical systems in the coalgebraic sense [6]. It is a requirement on a lossy projection, so invertibility is neither needed nor assumed.

Remark 4.2. The representation is loop-sensitive, since a decoration encodes the structure reachable from its node, cycles included. What the thema can lose is of three kinds. Nodes that are bisimilar in the associated membership graph, in which each element of a label counts as an additional child, receive the same decoration, so distinct graphs whose membership structure unfolds identically from the point share a thema. Bisimilarity here concerns membership structure, not the behavior of the network under the update S . The labels record only multiplicities of active incoming weights, so the identity of the active presynaptic neurons is not retained beyond those counts. And nothing is recorded about parts of the network that cannot be reached from the point along directed edges. A failure of closure therefore reflects missing information of these kinds, and not a failure to register cycles.

The passage from the full decorated network to the thema is where the lossless equivalences end. The foundations paper is explicit that the passage is many-to-one: many graphs have a given set as the decoration of their point, and the neural states sharing one thema are in principle unboundedly many [2]. There is a unique canonical tree whose decoration is a given set, so the thema determines its pointed graph once graphs are restricted to canonical trees, but not when arbitrary graphs are allowed. Under Aczel's theory two pointed graphs have the same decoration exactly when they are bisimilar [1]. For labeled graphs the relevant graph is the expanded one in which each element of a label is attached as an additional child, because the decoration equation merges label elements with successor decorations: a terminal node labeled $\{\emptyset\}$ and an unlabeled node with an edge to an unlabeled terminal node both receive the decoration $\{\emptyset\}$. The separation of well-founded labels from non-well-founded successor decorations in Section 2 rules out this particular ambiguity. The

authors list the study of bisimulation of graphs as future work. The dynamics paper calls a fixed pointed weighted multigraph special when, for every well-founded labeling, the decoration of the point uniquely determines the entire decoration on that graph, and it gives examples of special and non-special graphs [3, Definition 5.3 and Section 5.2]. This concerns recovery of the full decoration with the graph held fixed, and not reconstruction of arbitrary topology from a thema. Whenever distinct full decorations on a fixed graph share a thema, the same thematic state can correspond to different underlying states. This non-uniqueness does not by itself establish a failure of dynamical closure. Closure fails only if two such states yield different thematic states after an update, so the question is whether the distinctions lost in passing to the thema can affect its continuation, and not whether the full graph can be reconstructed.

A small example separates the two. Let full states be pairs (x, y) , let $\pi(x, y) = x$, and compare the updates

$$S_1(x, y) = (x, 0), \quad S_2(x, y) = (y, 0).$$

The projection is many-to-one in both cases. Under S_1 it is closed, with $F(x) = x$. Under S_2 it is not closed whenever different values of y occur with the same x , because $\pi S_2(x, y) = y$ is not a function of x . An injective projection is trivially closed, but injectivity is not necessary for closure.

This yields a question the paper leaves open and the closure condition makes precise. For which networks, and for which choices of distinguished node, is the projection to the thema closed under the dynamics? Where it is closed, the thematic level can be studied as a dynamical system in its own right. Where it is not, any description at that level is incomplete in a specific and measurable way, since the missing information is exactly what separates states that the viewpoint cannot tell apart but the future can.

5. THE CONSCIOUSNESS OPERATORS

The paper interprets a class of operators K as consciousness operators, subject to four axioms:

$$Kx \subseteq x, \quad x \notin Kx, \quad Kx \not\subseteq x, \quad x \subseteq y \implies Kx = x \cap Ky.$$

The authors name them Generation, Irreversibility, Removal, and Selection, and read them as conditions relating a primary experience x , modeled as a set, to the awareness Kx it induces [2]. Their semantic thesis is that consciousness is the result of such an operator acting on experience, and their central example is the Russell operator $\mathcal{R}x = \{y \in x \mid y \notin y\}$, which they show satisfies all four axioms. The axioms are definite mathematical restrictions, and it is worth seeing what they impose before asking what they mean.

The last axiom has a revealing consequence. Taking $\{a\} \subseteq x$ gives

$$K(\{a\}) = \{a\} \cap Kx,$$

so whether a is selected is settled by how the operator treats its singleton, independently of which larger set contains it. At this level K behaves as a fixed element-selection criterion. The foundations paper proves a stronger form of the point: an operator satisfying the fourth axiom is determined entirely by its action on singletons, through $Kx = \{y \in x \mid K\{y\} = \{y\}\}$, and two such operators compose by intersection, so that $K^2 = K$ and consciousness operators commute [2]. Applying a single operator a second time therefore changes nothing, and composing several yields their intersection. The selector W of Section 2 is an operator of the same kind, being selection by the predicate of being well-founded, and it satisfies all four axioms. The first is immediate, and the fourth holds because selection by any predicate satisfies $x \subseteq y \Rightarrow Wx = x \cap Wy$, as the foundations paper notes for operators of this form [2]. For the second, if $x \in Wx$, then x is well-founded, since W keeps only well-founded sets, and $x \in x$, which is impossible. For the third, if $Wx \in x$, then Wx , a set all of whose elements are well-founded, is itself well-founded and so belongs to Wx , which is again impossible. Relations among elements can already be encoded inside the elements themselves, but the operator does not vary its selection with the surrounding context. That makes it hard to identify the operator with attention or awareness, both of which are usually taken to depend on context, unless some further account of that dependence is supplied.

More generally, the axioms contain no independent condition that separates conscious from unconscious activity, no way to pick out a particular sensory quality, and no route to predicting a report of experience. The paper attaches semantic readings to its objects, and those readings are interpretations. Before the construction can count as an explanatory theory of consciousness it needs an empirical bridge, a specified correspondence between its objects and something observable, and the axioms do not provide one. The authors state that they do not claim a complete characterization of consciousness and present the approach as a way of introducing awareness into a theory [2], so the observation here is only that the axioms alone do not supply the bridge. The closure condition from Section 4 is the part of the framework that does not depend on the interpretation, since it can be stated and tested for any projection of any system.

6. SUBJECTIVITY AS A METHOD

A methodological move common in philosophy of mind defines the object of study as whatever is available from a point of view, and treats everything the viewpoint does not contain as outside the subject matter. In the vocabulary of this essay, a point of view is a projection. The thema is the paper's version of one: the state of the whole

network as seen from a single node. The move has a cost that is easy to overlook. If whatever the projection retains is called experience and whatever it discards is called non-experience, then no result can ever count against the identification, because every discrepancy is reclassified as lying outside the viewpoint. A claim protected in this way has no content as an empirical claim, however natural it feels as a description.

The closure condition restores content. Whether a first-person description is an autonomous level of explanation becomes a question with a determinate answer: do states that look the same from the viewpoint remain the same after the system updates? When $\pi S = F\pi$ holds, the viewpoint is dynamically closed, its states evolve according to a rule stated entirely in its own terms, and the subjective level can be studied on its own. When the condition fails, the same thematic state has different futures, and any description confined to the viewpoint depends on hidden information to continue. The refusal to answer for what lies outside a point of view, which philosophers adopt as a policy, is thereby turned into a statement about when the outside can safely be ignored.

This also clarifies why mathematical formulation has long served as a device for managing the difficulties of subjectivity. A structural result can be stated without reference to any individual, anecdote, or named case, and can be checked by anyone, which is what objectivity amounts to in practice: independence from who is making the claim. The qualification is that this covers the proofs and not the modeling choices that precede them. Deciding that the decoration at a node is the object to be called experience, or that one projection rather than another is the relevant one, is a choice made under assumptions and interests, and a formalism can hide such a choice by making its consequences look inevitable. A disciplined treatment therefore states the structural results in full generality and marks each point at which an interpretation is added, naming the node, the projection, and the label as separate decisions that each need their own defense.

7. SELECTION BY SUFFICIENCY

The closure condition applies to any mechanism that selects which features of a system are retained, not only to the particular projection studied above. Let σ be a selection that maps states to retained features, and let $\mathcal{T}$ be a family of operations that the retained features will be used to anticipate or perform. Say that σ is *sufficient* for $\mathcal{T}$ if it is closed under every operation in the family, in the sense of Section 4: for each $S \in \mathcal{T}$, states that agree under σ agree after S . A sufficient selection is one whose discarded distinctions cannot alter any outcome the family cares about. It is relative to $\mathcal{T}$, so a selection can be sufficient for one family of operations and insufficient for another, and the question of adequacy cannot be asked without naming the operations.

The requirement can also be relaxed to what is actually observed. If o is an output map naming what is read off after an operation, say that σ is sufficient for $(\mathcal{T}, o)$

when $\sigma(\mu_1) = \sigma(\mu_2)$ implies $o(S\mu_1) = o(S\mu_2)$ for every $S \in \mathcal{T}$. This asks that the observable result be reproduced without requiring that the retained features predict the entire subsequent state. Closure together with the requirement that o be determined by σ implies this condition, but the condition does not imply closure, since the retained features may fail to predict their own later values while still fixing the observed output.

A second kind of selection is governed by a ranking that does not refer to $\mathcal{T}$ at all, such as a ranking by salience that retains whatever has most strongly drawn attention. Nothing in the ranking itself guarantees closure under the operations of interest. A ranking can correlate reliably with task-relevant distinctions, through training, evolution, or environmental regularity, and then closure may hold, but that alignment is a further fact that needs its own support. A selection tuned to salience alone can discard exactly the distinctions that continuation depends on. The contrast, stated in these terms, is between selection that is established to be sufficient for the tasks at hand and selection that supplies no such guarantee, and it can be examined as a property of projections without reference to any particular system in which either occurs. Applied to the construction of Sections 1 to 4, the question is whether the thema projection is sufficient for the dynamics, which is the closure condition of Section 4 read as a sufficiency claim. The ranking case corresponds to the situation in which the projection is fixed by something other than the dynamics it will be used to predict, and sufficiency is then a further fact that needs its own support.

A recent case from machine learning shows the distinction operating in practice [8]. In sequential state tracking on five boxes, the state after k moves is an arrangement of the objects, and the arrangement after move $k + 1$ is fixed by the arrangement after move k together with that move. The arrangement is therefore a selection that is sufficient, in the sense above, for the family of single moves, although it discards nearly everything a hidden state of a language model could carry. Wang and colleagues train a weighted looped language model so that iteration k reports the arrangement after k moves, and they find that extrapolation to longer sequences depends on two conditions: each iteration must be supervised with the arrangement it is meant to hold, and the scale of the recurrent state must be held fixed so that the step map is the same function at every iteration. In their controlled ablations, removing the norm anchoring reduced extrapolation from sixteen times the training length to below about two and a quarter times, and replacing the arrangement targets with final-answer supervision reduced it to at most eight times. The reading offered here is this essay's and not the authors', who do not use its vocabulary. Supervising the final answer at every iteration asks an intermediate state to determine something it does not yet contain, and a drifting scale means that the projected update is no longer a single map F . The case also marks a difference from the construction studied above. For Miranker and Zuckerman the projection is fixed by the representation and closure is a fact to be tested, whereas in

the trained model the projection is shaped to be sufficient and closure is secured by training and not discovered. The results are those of a single preprint on synthetic tasks, and they illustrate the sufficiency condition without testing anything about the thema.

8. LIMITS

Several things the construction does not show should be stated plainly. It adds no causal mechanism to the underlying network. The set-based description evolves consistently with the threshold dynamics, as Section 3 argues, but a consistent redescription is not a new cause. It also does not establish an ontological division. The foundations paper separates physical neural states, which can be observed and measured, from the sets decorating their graphs, which it treats as Platonic and located in a virtual space. The mathematical distinction is real, but treating one side as physical and the other as virtual is an interpretation added to it, and a difference between mathematical representations does not establish a corresponding difference in what they represent. Circular sets can encode recurrence without recurrence thereby becoming experience.

There is also a computational qualification. The foundations paper observes that anti-foundation guarantees the existence of a decoration without giving a method for constructing it by recursion, and takes the resulting failure of effective construction to reinforce the Platonic status of the decorations [2]. This should be read narrowly, since what fails is the unfolding of a circular set into an explicit well-founded membership structure, and not the finite specification of the graph and its decoration. Failure of ordinary well-founded recursive evaluation does not imply that finite cyclic structures cannot be represented and manipulated by computation. A finite graph specifies a circular set without any algorithm having to unfold its full membership structure, and for finite graphs whose labels have finite membership representations, including the rational-weight histograms considered here, the question of whether two pointed graphs have the same decoration reduces to bisimilarity of the associated expanded membership graphs of Section 4, which is decidable by partition refinement [7]. The choice of anti-foundation therefore does not by itself demonstrate a capacity beyond finite graph processing, and the non-computability of the unfolded sets does not by itself support placing them outside the physical description.

Finally, the closure condition is a statement about a projection and a dynamics, and its verdicts are only as good as the model supplied. It says whether a viewpoint determines its own continuation within the model. Whether the model corresponds to anything beyond itself is a separate question that the formalism leaves open, and the open question about which networks and which distinguished nodes satisfy the condition remains the most direct place for the construction to be tested.

REFERENCES

- [1] P. Aczel, *Non-Well-Founded Sets*. CSLI Lecture Notes 14, CSLI Publications, Stanford, 1988.
- [2] W. L. Miranker and G. J. Zuckerman, "Mathematical foundations of consciousness," *Journal of Applied Logic*, vol. 7, no. 4, pp. 421–440, 2009. arXiv:0810.4339.
- [3] W. L. Miranker and G. J. Zuckerman, "Dynamics of mental activity," *Journal of Applied Logic*, vol. 8, no. 1, pp. 114–140, 2010. doi:10.1016/j.jal.2009.05.002.
- [4] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*. Van Nostrand, Princeton, 1960.
- [5] D. Park, "Concurrency and automata on infinite sequences," in *Theoretical Computer Science*, Lecture Notes in Computer Science, vol. 104, pp. 167–183. Springer, 1981.
- [6] J. J. M. M. Rutten, "Universal coalgebra: a theory of systems," *Theoretical Computer Science*, vol. 249, no. 1, pp. 3–80, 2000.
- [7] R. Paige and R. E. Tarjan, "Three partition refinement algorithms," *SIAM Journal on Computing*, vol. 16, no. 6, pp. 973–989, 1987.
- [8] P. Wang, Y. Hu, G. Dong, Y. Lin, H. Wang, Z. Liu, Z. Wen, J. Li, J.-M. Yang, W. Wu and Y. Tong, "PAR: Progress-aligned recurrence for length generalization in state tracking," arXiv:2610.27628, 2026.