Geometry from Histories

Categorical Reconstruction of Learned State Spaces

Victor Zhorin

Revised September 2026

Abstract

A state space reconstructed from histories must preserve the operations that give those histories their consequences. We formulate this requirement through a category of exact realizations of response laws and supported continuation. The complete response profile is terminal, and reduced exact realizations are therefore canonically isomorphic. A discounted operational metric gives a compact completion; finite-test approximation errors yield explicit Gromov–Hausdorff bounds with a separate term for unresolved continuation. Uniqueness also forces compatible local realizations to satisfy the cocycle condition. Temporal direction enters as an obstruction to representations that erase response-relevant order. Learned continuous-time signatures and attention provide a constructive realization whose compatibility can be examined under changes of recording and coordinates. ReLU components supply piecewise-affine local structure, with interface contributions retained by an exact distributional formula for finite contrasts. For systems with absorption, localization must intertwine killed continuation operators: this preserves termination laws, expected lifetime and transported responses, with quantitative bounds under approximate intertwining. Valuation and differential localization extend the construction when their additional operator hypotheses hold. The resulting framework reconstructs geometry through observable distinctions and specifies which dynamics and boundary effects must survive a change of representation.

1 Reconstructing a system from its histories

A sequence of observations does not arrive with its state space attached. The same history can be represented by a vector, a probability law or a point in a geometric space, and each representation permits different operations. The scientific question is which of those operations retain the consequences of the history. A geometry becomes meaningful when its states support the observations and continuations through which the system can be distinguished.

The basic requirement is compositional. Continue a history and then represent it, or represent the history and evolve the represented state: the two routes should agree. If they disagree, the representation has lost information needed for that continuation. This condition places dynamics inside the definition of an adequate state, before a metric, a manifold or a particular learning architecture is chosen.

The complete family of response laws determines a canonical object. Histories belong to the same state when no specified continuation experiment distinguishes them. We prove that their response profiles form the terminal object in the category of exact realizations. Every reduced realization is thus a coordinate description of the same observable system, with a uniquely determined comparison map. This universal property supplies the foundation for reconstruction and for compatibility between local descriptions.

Temporal direction belongs to this foundation. When differently ordered histories have different consequences, any representation that identifies them fails to preserve the response law. The distinction survives every exact change of realization. An operational arrow of time can therefore be expressed as a restriction on admissible representations, independently of whether the system has a terminal boundary.

The next question is quantitative: how much geometry is determined by finitely many observations? A weighted response metric measures distinguishability at a stated resolution. Its completion is compact for the finite operational language considered here, and uniform approximation on finite tests yields an explicit geometric error bound. The bound separates errors in resolved experiments from ambiguity in longer continuations. Increasing data and extending the observational horizon address different parts of the reconstruction problem.

Local observations introduce a further requirement. Different contexts may expose overlapping portions of one system without sharing coordinates. Where their profiles distinguish the same states, uniqueness determines the transition maps and forces the cocycle condition. This is the basis of localization across observation sources: the common object is recovered through compatible operations on its local realizations.

The computational construction follows the same logic. Learned continuous-time signatures govern the compatibility weights of attention, while ReLU components supply local piecewise-affine structure. Storage order, temporal order and observational restriction act differently on this construction. Their effects can be tested against the required commuting relations. The resulting geometry includes both smooth contextual variation and singular contributions at activation interfaces.

A terminal boundary strengthens the operator requirement. Absorption removes probability mass, and that loss determines termination probabilities and expected lifetime. It must survive localization together with the interior evolution. We show how killed-operator intertwining preserves these quantities and how approximate intertwining controls value error. Finite contrasts retain the associated response even when ordinary interior derivatives miss interface contributions or coordinates degenerate at the boundary.

This perspective continues the tensor-based programme in which computational representations are organized by the constraints and operators of a system [1]. Here, reconstruction makes the represented state itself an object of inquiry. The paper develops the general construction first; the research applications enter later as settings in which its preservation requirements can be examined.

1.1 Temporal direction as a reconstruction requirement

Let $B(h)$ retain event composition and any specified non-temporal marks while forgetting the temporal structure under investigation. Let $I_{c,\sigma}h$ insert a finite event chain c into h according to schedule σ. A directional comparison changes its orientation while preserving the retained information:

$$B(I_{c,\sigma}h) = B(I_{c,\sigma'}h). \tag{1}$$

The response contrast is

$$A_r(h; c, \sigma, \sigma') = r(I_{c,\sigma}h) - r(I_{c,\sigma'}h). \tag{2}$$

The comparison specifies which marks and time intervals remain fixed. Reversing orientation, changing spacing and translating a time origin are distinct operations; the chosen contrast identifies which temporal distinction the response requires.

Proposition 1.1 (Obstruction to forgetting temporal structure)**.** *Suppose $B(h) = B(h')$ and $|r(h) - r(h')| = \Delta > 0$. Every readout g depending only on B satisfies*

$$\max\{|r(h) - g(B(h))|, |r(h') - g(B(h'))|\} \geq \Delta/2. \tag{3}$$

Consequently r cannot factor exactly through B.

Proof. Set $z = g(B(h)) = g(B(h'))$. The triangle inequality gives $\Delta \leq |r(h) - z| + |z - r(h')|$. □

A nonzero contrast makes temporal orientation an unavoidable distinction for the chosen response law. Every exact realization must preserve it, regardless of the coordinates used. This is the operational arrow of time in the reconstruction problem. Its scientific interpretation follows the law being represented: a learned response establishes the model's distinction, and an identified causal response establishes the corresponding causal distinction. Neither construction requires an absorbing boundary.

The category introduced below gives this requirement its full form by retaining a family of responses together with their supported continuations. The boundary will subsequently enlarge the structure that the same comparison maps must preserve.

2 A worked reconstruction: one score, different futures

Consider a population of machines with two persistent latent modes $Z \in \{0, 1\}$. A history determines an information state $\theta = \mathbb{P}(Z = 1 \mid h)$. The example specifies three available one-step experiments:

Experiment and reported outcome	$Z = 0$	$Z = 1$
Baseline operation: success	0.2	0.2
Diagnostic measurement: positive	0.1	0.9
Maintenance attempt: success	0.9	0.1

Conditional on the fixed mode, successive experimental outcomes are independent in this illustrative model. A maintenance attempt need not change the mode, which can represent an immutable failure type. Its success is an outcome with the stated mode dependence. The response probabilities are

$$b(\theta) = 0.2, \qquad d(\theta) = 0.1 + 0.8\theta, \qquad m(\theta) = 0.9 - 0.8\theta. \tag{4}$$

A model trained only on baseline success can be perfectly calibrated with a constant state. Every θ is equivalent for that task. A diagnostic experiment separates them. For any full test t, the law has the mixture form

$$p_\theta(t) = (1 - \theta)p_0(t) + \theta p_1(t).$$

Thus one scalar suffices for all experiments in this model, although a scalar baseline risk does not. Dimension alone says nothing about whether the right distinction has been retained.

On the three one-step tests, the unweighted response metric is

$$d_{\mathrm{one}}(\theta, \theta') = 0.8|\theta - \theta'|. \tag{5}$$

If population contexts supply a dense set of posterior values, the completion is the interval $[0, 1]$. This continuum describes distinguishable information states of a system with only two latent modes. It is an effective geometry, not evidence that the physical mode itself is continuous.

A positive diagnostic updates the state by

$$\theta^+ = \frac{0.9\theta}{0.1 + 0.8\theta}, \qquad \frac{\mathrm{d}\theta^+}{\mathrm{d}\theta} = \frac{0.09}{(0.1 + 0.8\theta)^2}. \tag{6}$$

For $\theta = 1/4$, the posterior becomes $3/4$. The recorded measurement changes the information state while the latent mode remains fixed. With repeated positive and negative observations the odds multiply by 9 and $1/9$, respectively. The update composes exactly, supplying an operational interpretation of motion in this state space.

Suppose a maintenance attempt costs 0.3 units and success yields one unit. Baseline operation has value 0.2, while maintenance has value $0.6 - 0.8\theta$. Hence

$$V(\theta) = \max\{0.2,\, 0.6 - 0.8\theta\}. \tag{7}$$

The preferred action changes at $\theta = 1/2$. Histories with the same baseline prediction require different decisions. The kink in V arises from optimization over smooth response branches; it is distinct from both a physical discontinuity and the information update in (6).

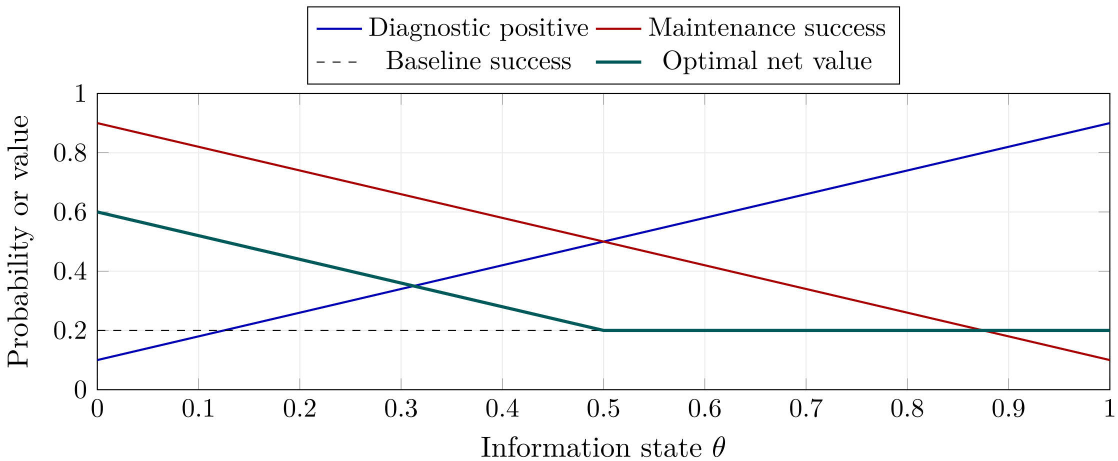

Figure 1: Analytic example, not empirical data. One outcome collapses the entire state space; additional experiments reveal a continuous information coordinate and an action threshold.

The example isolates the scientific issue. A successful predictor may already contain enough information for other tasks, but its accuracy on one target does not establish that fact. The relevant question is whether additional response functions factor through the learned state. In a real application, this can be investigated without discarding the predictor or reconstructing its internal architecture.

The example will recur in the general results. Its diagnostic readout supplies a separating coordinate, its Bayes update supplies a commuting continuation map, and its action threshold supplies a genuine kink in a value function. Each structure is explicit enough to distinguish a reconstruction claim from its assumptions. In particular, adding diagnostic observations changes the operational quotient even though the baseline risk remains constant.

3 Experiments, histories, and the observable state

3.1 A finite operational language

Fix finite sets of actions $\mathcal{A}$ and observations $\mathcal{O}$. An action can represent an intervention, a measurement setting, or a prescribed waiting interval. An uncontrolled process is included by taking a single action. A history is a finite word

$$h = a_1 o_1 \cdots a_k o_k.$$

Let $\mathcal{H}$ contain the reachable histories of a specified controlled process, including any initial contexts being compared. We assume all actions are available; state-dependent feasibility can instead be encoded as an observable response or handled with typed domains.

A test $t \in \mathcal{T} = (\mathcal{A} \times \mathcal{O})^*$ is a finite future word. Its response

$$p_h(t) = \mathbb{P}(o_1, \ldots, o_m \mid h;\ a_1, \ldots, a_m \text{ prescribed}) \tag{8}$$

is the probability of its observation sequence when its actions are imposed in sequence. The semicolon distinguishes a prescribed protocol from conditioning on action choices in an observational dataset. These probabilities are defined by a controlled law. Their identification from any particular dataset is addressed in Section 13.

Set $p_h(\varnothing) = 1$. Sequential consistency gives

$$p_h(ao\,t) = p_h(ao)\,p_{hao}(t), \qquad p_h(ao) > 0, \tag{9}$$

and $\sum_o p_h(tao) = p_h(t)$ for every test t and next action a. The complete family of tests includes arbitrarily long continuations. Rewards can be included in observations when they are observable, or added as a separate family of required readouts.

This language is the operational basis of predictive-state representations. Equality of future laws also underlies the minimal predictive states of computational mechanics [6, 7]. The state construction below makes their role explicit in a reconstruction problem. Categorical treatments of behavioral distances provide a broader setting in which both trace and branching semantics can be studied [8].

3.2 The canonical response profile

Definition 3.1 (Predictive equivalence)**.** Histories h, h' are equivalent if $p_h(t) = p_{h'}(t)$ for every $t \in \mathcal{T}$. Write $h \sim h'$, and define

$$q(h) = (p_h(t))_{t\in\mathcal{T}}, \qquad Q = q(\mathcal{H}) \subseteq [0,1]^{\mathcal{T}}. \tag{10}$$

Thus Q is the set of predictive equivalence classes expressed as response profiles.

In the example of Section 2, the profile includes diagnostic and maintenance responses as well as baseline success. The profile is not a list of predicted labels from one task. It describes the system's observable behavior under continuation. A single risk score is one coordinate or one function of this object. It can identify fewer distinctions than the state required for further observation or control.

Proposition 3.2 (Closed predictive update)**.** *For $q \in Q$ with $q(ao) > 0$, the rule*

$$F_{ao}(q)(t) = \frac{q(ao\,t)}{q(ao)} \tag{11}$$

defines a unique next profile, and $F_{ao}(q(h)) = q(hao)$. The distribution of next profiles under action a is

$$\overline{K}_a(q,\cdot) = \sum_{o:q(ao)>0} q(ao)\,\delta_{F_{ao}(q)}. \tag{12}$$

Proof. If two histories have the same profile, both numerator and denominator in (11) agree. Equation (9) identifies the ratio with the next-history test probability. Summing the one-step probabilities gives one, establishing (12). Zero-probability branches carry no transition mass and require no conditional state. □

The update is the familiar predictive-state conditioning identity [7]. Its importance here is structural: an equivalence defined by all future tests survives every supported observation. A partition defined by one present score need not have this property. The Bayes update in (6) is an explicit instance. No smooth manifold is used in this argument.

For real-valued timestamps or marks, the finite alphabet represents a stated observational resolution, not a claim that physical time is discrete. Standard Borel observation spaces require kernels, determining families of measurable tests, and almost-everywhere conditional updates. The finite-alphabet results below are stated at their actual level of generality; continuous-time extensions need additional regularity and coverage assumptions.

3.3 A category of exact realizations

Definition 3.3 (Realization)**.** An exact realization of the process consists of a surjection $\phi : \mathcal{H} \to X$, readouts $r_t : X \to [0,1]$, and supported updates G_{ao} satisfying

$$r_t\phi = p_{(\cdot)}(t), \qquad G_{ao}\phi(h) = \phi(hao). \tag{13}$$

A morphism from (X, ϕ, r, G) to (Y, ψ, s, H) is a map $J : X \to Y$ satisfying

$$J\phi = \psi, \qquad s_t J = r_t, \qquad JG_{ao} = H_{ao}J \tag{14}$$

on supported domains. A realization is reduced if its readouts separate its states.

Composition of these maps is ordinary composition, and the identities satisfy the same equations. This defines a category $\mathsf{Real}(p)$ for the fixed response law p. Surjectivity excludes unused states whose behavior is unconstrained by the histories under consideration.

Theorem 3.4 (Minimal reconstruction)**.** *The profile realization Q is terminal in $\mathsf{Real}(p)$. Every reduced exact realization is uniquely isomorphic to Q through a map commuting with the history representation. Consequently any two reduced exact realizations are uniquely isomorphic in this sense.*

Proof. For a realization X, define $J(x) = (r_t(x))_t$. Every x equals $\phi(h)$ for some h, so $J(x) = q(h) \in Q$. If $\phi(h) = \phi(h')$, exactness makes their profiles equal, hence the definition is consistent. Readout preservation is immediate, and update preservation follows from (9). The equation $J\phi = q$ determines J uniquely. It is surjective. If X is reduced, equal profiles imply equal states, so J is injective; its inverse preserves the readouts and supported updates as well. $\square$

The theorem is a universal-property formulation of predictive minimality. It states what is common to exact realizations, not that learning has already recovered one. Category theory is doing a specific job: specifying which changes of representation preserve the scientific content, and identifying the unique minimal object relative to that content.

The commuting relation can be displayed as

$$\begin{tikzcd}
\mathcal{H} \arrow[r, "h\mapsto hao"] \arrow[d, "\phi"'] & \mathcal{H} \arrow[d, "\phi"] \\
X \arrow[r, "G_{ao}"'] \arrow[d, "J"'] & X \arrow[d, "J"] \\
Q \arrow[r, "F_{ao}"'] & Q.
\end{tikzcd}$$

Each square is a testable preservation requirement on its supported domain. Treating a vector as a state does not make the square commute.

If only baseline success is retained in the example, every θ maps to one profile. Adding the diagnostic test makes $\theta \mapsto 0.1 + 0.8\theta$ injective. The theorem formalizes this distinction through separating readouts rather than through the dimension of the latent vector.

4 From observable equivalence to operational geometry

The minimal reconstruction theorem supplies a state set and update laws. It does not supply a Euclidean metric. An arbitrary injection of Q into a vector space can stretch some regions and compress others without changing any test response. Geometry becomes identifiable only after specifying how response differences are to be compared.

4.1 A metric with explicit resolution

Choose $0 < \gamma < 1$ and let $|t|$ be the number of action–observation pairs in a test. Define

$$d_\gamma(q, q') = \sup_{t \in \mathcal{T}} \gamma^{|t|} |q(t) - q'(t)|. \tag{15}$$

The parameter γ determines the weight assigned to increasingly long experiments. It is part of the observational specification, not a learned law of nature. If elapsed time rather than event length is the relevant scale, another weight can be chosen; the compactness argument below then requires a suitable finite approximation property.

Theorem 4.1 (Observable metric and compact completion)**.** *For finite $\mathcal{A}, \mathcal{O}$, d_γ is a metric on Q. It induces the product topology inherited from $[0,1]^{\mathcal{T}}$. Its completion $\widehat{Q}$ is the closure of Q in that product, and is compact. Every point of $\widehat{Q}$ satisfies the finite-test probability consistency identities.*

Proof. Symmetry and nonnegativity are immediate. The triangle inequality holds coordinatewise and is preserved by taking a supremum. Zero distance means equality of every coordinate, hence equality of profiles. For $L \geq 0$, put

$$d_{\gamma,L}(q, q') = \max_{|t| \leq L} \gamma^{|t|} |q(t) - q'(t)|.$$

There are finitely many tests of length at most L, and

$$d_{\gamma,L} \leq d_\gamma \leq \max\{d_{\gamma,L}, \gamma^{L+1}\}. \tag{16}$$

Coordinatewise convergence therefore implies metric convergence by controlling a finite set of coordinates and then the tail. Conversely, $|q(t) - q'(t)| \leq \gamma^{-|t|} d_\gamma(q, q')$, so metric convergence implies coordinatewise convergence. The same argument metrizes the full product, which is compact. Its closed subset $\widehat{Q}$ is compact and complete, and contains Q densely. Probability normalization and consistency are closed finite-coordinate equalities, so they hold on the closure. □

This is a precise form of emergence from discrete histories. A countable collection of finite histories can have a compact, uncountable completion in the topology of observable continuation. The result does not force this completion to be a manifold. It can contain disconnected components, singular strata, or infinitely many effective degrees of freedom. What emerges automatically is an operational topological space. Manifold structure is a further empirical and mathematical claim.

The completion also has a useful interpretation. Its additional points are limits of distinguishable histories at progressively finer observational resolution. They need not correspond to histories that occur with positive probability in a fixed finite dataset. Conflating completion with observed support would turn a mathematical construction into an unsupported population claim.

Proposition 4.2 (Regularity of supported updates)**.** *On the set where $q(ao), q'(ao) \geq \eta > 0$,*

$$d_\gamma(F_{ao}q, F_{ao}q') \leq \frac{2}{\eta\gamma}\, d_\gamma(q, q'). \tag{17}$$

The same update extends to the positive-probability domain of $\widehat{Q}$ and takes values in $\widehat{Q}$.

Proof. Write $u = q(ao\,t)$, $v = q'(ao\,t)$, $p = q(ao)$, and $p' = q'(ao)$. Since $v \le p'$,

$$\left|\frac{u}{p} - \frac{v}{p'}\right| \le \frac{|u-v| + |p-p'|}{p}.$$

After multiplication by $\gamma^{|t|}$, both numerator terms are bounded by $\gamma^{-1} d_\gamma(q,q')$. Taking the supremum proves (17). For a sequence of reachable profiles tending to a limit with $q(ao) > 0$, the branch probabilities are eventually bounded below. Their updated profiles converge by the same bound and have the ratio coordinates (11); thus the limit belongs to the closure. □

The dependence on η is informative. Conditioning on a rare observation can magnify small predictive differences. A sharp change in an information state can therefore arise from ordinary conditional probability. Such a change need not represent an instantaneous change in the underlying physical system.

4.2 Uniqueness up to observable isometry

On a reduced exact realization X, equip the state space with

$$d_X(x,x') = \sup_t \gamma^{|t|} |r_t(x) - r_t(x')|. \tag{18}$$

Corollary 4.3 (Geometric reconstruction)**.** *The unique isomorphism between two reduced exact realizations is an isometry for their metrics* (18)*. It extends uniquely to an isometry of their completions.*

Proof. The isomorphism preserves every readout, hence every term in the supremum. An isometry between dense subspaces extends uniquely to their completions. □

The metric is reconstructed because its defining readouts are preserved. In the worked example, the three one-step tests recover the interval metric in (5); the baseline test alone gives zero distance everywhere. The conclusion concerns the metric selected by the experimental family, not an unrelated Euclidean metric attached to a realization.

4.3 Finite observations and quantitative ambiguity

An empirical model supplies estimates on a finite family of tests. Let $\mathcal{H}_0$ be a nonempty set of histories being compared. For two encoders ϕ_i with decoded predictions $\widetilde{p}_i(h,t) = r_t^i(\phi_i(h)) \in [0,1]$, suppose

$$\sup_{h\in\mathcal{H}_0,\ |t|\le L} |\widetilde{p}_i(h,t) - p_h(t)| \le \varepsilon_i, \qquad i = 1,2. \tag{19}$$

Define $\widetilde{d}_{i,L}$ by (15) using these decoded predictions and restricting to $|t| \le L$.

Theorem 4.4 (Finite-resolution reconstruction bound)**.** *Under* (19)*, for all* $h, h' \in \mathcal{H}_0$*,*

$$|\widetilde{d}_{i,L}(h,h') - d_\gamma(q(h), q(h'))| \le 2\varepsilon_i + \gamma^{L+1}, \tag{20}$$

$$|\widetilde{d}_{1,L}(h,h') - \widetilde{d}_{2,L}(h,h')| \le 2(\varepsilon_1 + \varepsilon_2). \tag{21}$$

If $\mathcal{H}_0$ *is finite, quotienting zero-distance pairs produces two finite metric spaces with Gromov–Hausdorff distance at most* $\varepsilon_1 + \varepsilon_2$*.*

Proof. For each test, the error in a difference of two predictions is at most $2\varepsilon_i$. Absolute values and finite maxima are nonexpansive in the supremum norm, so $|\widetilde{d}_{i,L} - d_{\gamma,L}| \le 2\varepsilon_i$. Equation (16) proves (20); the triangle inequality proves (21). Pair the two quotient points represented by each h. This is a correspondence onto both finite spaces, with distortion bounded by $2(\varepsilon_1 + \varepsilon_2)$. The correspondence characterization of Gromov–Hausdorff distance gives half this bound. □

The last assertion can also be obtained by constructing a common metric space from the correspondence, assigning a cross-space separation slightly larger than half its distortion, and taking an infimum. Thus the geometric comparison depends on response accuracy and shared coverage. It does not require that the raw latent vectors have the same dimension.

The horizon term in (20) represents unresolved experiments. The estimation terms represent errors in the resolved experiments. These are different sources of ambiguity. Increasing the number of samples at a fixed horizon cannot remove the first, while increasing the horizon without adequate samples can worsen the second.

The theorem applies to the specified set $\mathcal{H}_0$. A held-out finite sample does not establish a uniform bound on an entire population. Nor does the theorem bound the Euclidean distance between latent vectors. That requires an additional comparison between the model's ambient metric and its decoded operational metric.

4.4 What the bound resolves

The comparison between two reconstructions and the comparison with the complete process are different quantities. The former uses their common resolved tests and has bound $\varepsilon_1 + \varepsilon_2$ on the finite quotient spaces. The latter also has the unresolved-horizon term γ^{L+1}. Agreement between two models on a short test family can therefore coexist with a common failure to distinguish longer continuations.

For example, at $\gamma = 0.9$, $L = 20$ and $\varepsilon_1 = \varepsilon_2 = 0.01$, the two reconstructed quotient spaces have Gromov–Hausdorff distance at most 0.02. Yet the bound for either decoded metric against the complete operational metric is approximately 0.1294, because $0.9^{21} \simeq 0.1094$. Increasing the horizon to 60, with the same assumed error bound, reduces that latter bound to approximately 0.0216. These are analytic substitutions, not estimates from data.

The history–test correspondence turns response accuracy into a geometric error bound. Boundary-sensitive readouts introduce a further scale dependence in Section 8: resolving a finite response requires accuracy at the scale of the perturbation.

5 Local realizations and the cocycle forced by uniqueness

A common represented system may be accessible only through local observation contexts. The key issue is whether their descriptions refer to the same distinguishable continuations. This can be formulated before selecting a geometric coordinate atlas.

Let Q be the profile space and let $U_i \subseteq Q$ cover the portion being reconstructed. Suppose each local realization has a bijective profile map $r_i : X_i \to U_i$. This means that the local observations distinguish every profile in the stated domain. Define the overlap in X_i by $r_i^{-1}(U_i \cap U_j)$ and the comparison by

$$J_{ij} = r_j^{-1} r_i. \tag{22}$$

For local continuation operators, impose compatibility wherever both a state and its successor belong to the relevant domains. The comparisons are then restrictions of the canonical realization isomorphisms.

Theorem 5.1 (Canonical descent)**.** *On their common domains the maps* (22) *satisfy*

$$J_{ii} = \mathrm{id}, \qquad J_{ji} = J_{ij}^{-1}, \qquad J_{jk}J_{ij} = J_{ik}. \tag{23}$$

Identifying overlap points by these maps reconstructs the covered profile space as a set. If the U_i form an open cover in the operational topology and the r_i are homeomorphisms, the same quotient construction recovers that topology.

Proof. Each equality follows by cancellation of the profile maps. Equivalently, the two paths across a triple overlap preserve the same history profile, so Theorem 3.4 makes them equal. Define the map from the disjoint union of X_i to $\bigcup_i U_i$ by $x \mapsto r_i(x)$. It is surjective and identifies exactly the specified overlap points, proving the set statement. Under the topological hypotheses it is continuous and open, since each r_i maps open subsets to sets open in the open subset U_i of Q. It therefore induces a homeomorphism from the quotient. □

The cocycle condition is a consequence of a common represented object. It does not have to be added as an independent numerical alignment constraint. This is the categorical dividend of the universal property: local uniqueness gives path-independent comparison. It applies when the local readouts separate the same profiles, not merely when each institution has a good predictor for its own target.

The qualification can be seen in the worked example. A context observing diagnostics has an injective profile coordinate. A context observing only baseline success collapses every θ. There is no invertible comparison between these contexts on the full interval. It would be incorrect to infer one from the fact that both baseline predictors are perfectly calibrated. Adding a common separating experiment restores the missing comparison.

5.1 Compatibility errors have observable consequences

For approximate local readout maps $\widetilde{r}_i : X_i \to Z$ into a common metric profile space (Z, d), suppose proposed transports satisfy

$$d(\widetilde{r}_j(J_{ij}x), \widetilde{r}_i(x)) \le \varepsilon_{ij}$$

on the relevant domains. Then the two routes on a triple overlap obey

$$d(\widetilde{r}_k(J_{jk}J_{ij}x), \widetilde{r}_k(J_{ik}x)) \le \varepsilon_{jk} + \varepsilon_{ij} + \varepsilon_{ik}. \tag{24}$$

This follows from the triangle inequality through the intermediate decoded profiles. It bounds an operational coherence defect. A bound in raw latent coordinates additionally needs a stable inverse to the readout map.

Agreement of probability marginals on overlaps is weaker. Fair binary pair laws with $A = B$, $B = C$ and $A \neq C$ have matching one-variable marginals but admit no global joint law. Even compatible marginals can admit multiple joint laws. Such obstructions explain why the hypothesis of a common separated profile object cannot be replaced by pairwise data agreement [9]. The attention construction below gives a further, architectural source of restriction defects.

6 When an operational space admits smooth geometry

A compact metric completion is already a geometry, but it is weaker than a differentiable manifold. Smooth structure becomes useful when a finite family of observations provides regular coordinates and when changes in those coordinates support meaningful sensitivity calculations.

6.1 Separating coordinates and regularity

Let M be a compact Hausdorff candidate realization, and let $r_1, \ldots, r_m$ be continuous readouts. If

$$R : M \to \mathbb{R}^m, \qquad R(x) = (r_1(x), \ldots, r_m(x))$$

is injective, it is a homeomorphism onto its image: a continuous bijection from a compact space to a Hausdorff space has continuous inverse. Thus a finite separating family determines the topology of this realization.

If M is already a smooth d-manifold, R is smooth and injective, and DR has rank d everywhere, compactness makes R a smooth embedding. These hypotheses explain what a smooth reconstruction must establish. They do not deduce smoothness from a cloud of points. The local rank condition distinguishes observable directions; global injectivity excludes distant states with identical measured responses.

Theorem 6.1 (Smooth reconstruction from separating experiments)**.** *Let M and N be compact smooth manifolds, possibly with boundary, containing reduced exact reachable realizations of the same history–test law densely. Assume all test readouts extend continuously and supported updates extend continuously on their positive-probability domains. Suppose a common finite collection of test readouts defines smooth embeddings*

$$R_M : M \to \mathbb{R}^m, \qquad R_N : N \to \mathbb{R}^m.$$

Then the unique isomorphism of the reachable realizations extends to a diffeomorphism $J : M \to N$. Every tensor constructed by pulling back the same smooth tensor on the common response image is preserved by J. In particular, a nondegenerate Fisher metric for a common smooth experimental family and fixed design is preserved.

Proof. Equality of test probabilities gives the same image for the reachable histories under R_M and R_N. Density, continuity, and compactness give $R_M(M) = R_N(N)$ by taking closures. Since both maps are smooth embeddings, each is a diffeomorphism onto this common embedded submanifold. Thus $R_N^{-1} R_M$ is a diffeomorphism extending the unique history-preserving map. Equality of all extended readouts and compatibility of supported updates follow by continuity from the dense reachable subsets. For a tensor A on the response image,

$$J^*(R_N^* A) = (R_N J)^* A = R_M^* A.$$

The Fisher statement follows directly from equality of the experimental probabilities under J and the chain rule, using formula (26) below. □

The theorem upgrades operational equivalence to equivalence of smooth realizations when finitely many experiments provide regular separating coordinates. Its force lies in making the extra hypotheses visible: smoothness, global separation, and regularity of the observation maps. Its conclusion does not extend to an unrelated metric attached to either realization. The scientific task is to establish a family of experiments for which these hypotheses hold, approximately or exactly.

Different experimental families can give the same topology with different metrics. In particular, every d_γ of Section 4 induces the same product topology for finite alphabets, while assigning different weights to long tests. This is an example of robust topological structure with nonunique metric scale. A claim about geometry should specify which of these levels is intended.

6.2 Reachable closure and coverage

Density in an operational completion is automatic; density in a proposed manifold is a separate assertion. Write $M_{\text{reach}} = \overline{\phi(\mathcal{H})}^{M}$. If this is a proper closed stratum of M, the response data reconstruct at most M_{reach}. The smooth theorem applies to that stratum only if it satisfies the stated smoothness and separating-coordinate hypotheses. There is no justified extension to the unobserved remainder of M.

A conditional sampling criterion makes the extra burden explicit. Suppose a candidate compact metric space M has an $\varepsilon/2$-net of N_ε centers and the sampling law gives each corresponding open ball mass at least $p_\varepsilon > 0$. For n independent sampled states,

$$\mathbb{P}(\text{the sample is not an } \varepsilon\text{-net of } M) \le N_\varepsilon(1-p_\varepsilon)^n \le N_\varepsilon e^{-np_\varepsilon}. \tag{25}$$

Indeed, if every ball contains a sampled point, each point of M lies within ε of a sample; the union bound controls the missed balls. Full support of an iid sampling law on a compact metric space similarly implies almost-sure density of an infinite sample, by applying this argument at a countable sequence of resolutions.

Equation (25) is a verification criterion conditional on a justified coverage model. A finite point cloud cannot establish positive mass in every unobserved neighbourhood. Repeated histories from one dependent trajectory also require dependence-aware sampling bounds. The theorem thus identifies both the reconstruction on observed support and the additional assumption required to identify it with a larger manifold.

6.3 A differential metric from experiments

Suppose experiment e has a finite outcome distribution $p_e(y \mid x) > 0$, smooth in $x \in M$. Choose fixed experimental weights $w_e > 0$ with $\sum_e w_e = 1$ over a finite family. The Fisher quadratic form is

$$g_x(v,v) = \sum_e w_e \sum_y p_e(y \mid x) \left(\mathrm{d}_x \log p_e(y \mid x)[v]\right)^2. \tag{26}$$

It measures local distinguishability under the specified experiments. It is positive definite precisely when no nonzero tangent direction leaves all these outcome probabilities unchanged to first order. Otherwise it is a semidefinite form and the unresolved tangent directions remain visible as its null space.

Proposition 6.2 (Coordinate invariance and loss under coarsening)**.** *Under a smooth change of coordinates,* (26) *transforms as a covariant two-tensor. If an experimental outcome is passed through a stochastic channel independent of x, its Fisher quadratic form cannot increase.*

Proof. Coordinate invariance follows by applying the chain rule to the differentials in (26). For coarsening, let Z be the output of a fixed channel applied to Y, and let $s_v(Y) = \mathrm{d}\log p(Y \mid x)[v]$. Differentiating the finite sum for $p(Z \mid x)$ shows that the coarsened score is $\mathbb{E}[s_v(Y) \mid Z]$. Jensen's inequality gives $\mathbb{E}[\mathbb{E}(s_v \mid Z)^2] \le \mathbb{E}[s_v^2]$. $\square$

For the diagnostic experiment in Section 2,

$$g(\theta) = \frac{0.64}{(0.1+0.8\theta)(0.9-0.8\theta)}. \tag{27}$$

The associated one-dimensional distance is

$$\ell(\theta_0,\theta_1) = 2\left|\arcsin\sqrt{0.1+0.8\theta_1} - \arcsin\sqrt{0.1+0.8\theta_0}\right|.$$

This metric and (5) describe the same interval but answer different comparison questions. The baseline experiment alone has zero Fisher information about θ, because its probability is constant. Adding a separating observation creates measurable local geometry.

The experimental design remains part of the metric. Changing w_e changes which distinctions are emphasized. If the design is state-dependent, one must decide whether the selection of an experiment is itself an observed random variable carrying information. Formula (26) uses a fixed external design and avoids mixing those two questions.

7 Temporal attention as constructive gluing

To realize the operational construction with irregular time, write a typed history as

$$h = ((e_1, t_1, m_1, d_1), \ldots, (e_n, t_n, m_n, d_n)), \qquad t_1 \le \cdots \le t_n. \tag{28}$$

Here e_i is an event, t_i its time, m_i its additional marks and d_i its domain. The finite operational language of Section 3 specifies observational resolution; the computational realization can use continuous timestamps. A Time2Vec map has the form

$$\tau(t) = \big(\omega_0 t + \varphi_0, \sin(\omega_1 t + \varphi_1), \ldots, \sin(\omega_k t + \varphi_k)\big). \tag{29}$$

Its frequencies and phases are learned. Cosine components can be included as additional phase-shifted sine components. These are temporal feature signatures; they are distinct from the iterated-integral signature of a path. Time2Vec supplies a model-independent representation of time that can be combined with event and domain embeddings [10].

Write the resulting token as

$$u_i^{(b)} = E(e_i) + D(d_i) + C(m_i) + bB_{\mathrm{time}}\tau(t_i), \qquad b \in \{0, 1\}. \tag{30}$$

The map C encodes the additional event marks. The switch b isolates the sequence-time contribution, and below $u_i = u_i^{(1)}$. The sum is a convenient notation for a common feature space; concatenation followed by a projection gives another implementation. No ordinal position feature is introduced. For one attention head,

$$q_i = W_Q u_i, \quad k_i = W_K u_i, \quad v_i = W_V u_i, \qquad \alpha_{ij} = \frac{\exp(s_{ij})}{\sum_{r \in I_i} \exp(s_{ir})}, \quad s_{ij} = \frac{q_i^\top k_j}{\sqrt{d_k}}, \tag{31}$$

where I_i is the set allowed by the mask. The contextual representation is

$$a_i = \sum_{j \in I_i} \alpha_{ij} v_j. \tag{32}$$

Multihead attention and residual connections combine several such operations. These formulas specialize the attention construction [11] to temporal signatures rather than sequence positions.

The gluing is constructive: a local token becomes meaningful through its coupling to other events in the history. Both the event content and the temporal signature contribute to the compatibility weights. At later layers, the same operation couples already contextualized representations. It is a learned assembly process whose coefficients depend on the object being assembled.

Proposition 7.1 (Independence of storage order)**.** *At inference, consider tokenwise shared maps and attention without ordinal position features. If a permutation P reorders the event tokens together with their timestamps, marks, domains and mask, then the contextual token array is permuted by P. A fixed CLS token has an invariant output when the permutation fixes its index.*

Proof. The score matrix transforms as $S \mapsto PSP^\top$. Rowwise masked softmax gives $A \mapsto PAP^\top$, and the value array transforms as $V \mapsto PV$. Thus $AV \mapsto PAV$. Shared tokenwise maps, residual addition and tokenwise normalization preserve this relation, so it holds through all layers. □

With $b = 0$ and a permutation-invariant mask and pooling rule, the same argument gives an exact bag-of-augmented-events model: retained marks and domain features remain attached to events. The exact symmetry is a consequence of these architectural conditions.

This invariance concerns the storage order of unchanged events. Reassigning timestamps changes the history and generally changes the output. An index-based causal mask must also be transported under reindexing; otherwise the mask itself introduces a different order. A chronology-based mask can be defined directly from event times. An encoder mask allowing all observed history tokens need not impose a causal order on its attention edges.

7.1 Which neighborhood belongs to the history?

Storage-order equivariance distinguishes two different transformations. Reindexing events together with their metadata changes their representation in memory. Reassigning their timestamps changes the temporal history. An architecture can respect the first transformation while remaining sensitive to the second.

The attention mask belongs to this distinction. Let $M(h)$ be the allowed-edge matrix. Equivariance requires

$$M(Ph) = PM(h)P^\top. \tag{33}$$

A mask defined from transported event times and marks can satisfy this relation. A fixed band defined by array indices generally does not when recomputed after reindexing. Thus temporal features inside the score cannot, by themselves, remove an ordinal restriction imposed on the available interactions.

This matters under changes in recording density. Inserting records between two retained events changes their token distance while preserving their elapsed-time separation. An index-window rule may remove their direct interaction. A rule based on elapsed time preserves that pair's eligibility, although the added information can still alter its attention weight through normalization. These are distinct effects: one changes which interactions are available, and the other changes their contextual weighting.

For cross-source reconstruction, only specified changes of recording should be identified. Additional informative events can legitimately change the state. Where a recording transformation C preserves the chosen operational law, an exact encoder must satisfy $J_C\phi(h) = \phi'(Ch)$ for the corresponding realization comparison. A failure of this relation measures dependence on the recording beyond the distinctions required by that law. This is the architectural form of the commuting requirement in Theorem 3.4.

7.2 The temporal part of the operator

Because τ is differentiable, temporal sensitivity can be computed explicitly:

$$\tau_j'(t) = \omega_j \cos(\omega_j t + \varphi_j), \qquad \tau_j''(t) = -\omega_j^2 \sin(\omega_j t + \varphi_j), \quad j \geq 1. \tag{34}$$

For a scalar perturbation parameter r on a fixed mask domain,

$$\partial_r \alpha_{ij} = \alpha_{ij} \left(\partial_r s_{ij} - \sum_k \alpha_{ik} \partial_r s_{ik} \right). \tag{35}$$

Hence

$$\partial_r a_i = \sum_j \alpha_{ij} \partial_r v_j + \sum_j (\partial_r \alpha_{ij}) v_j. \tag{36}$$

The second term records a change in which events the representation emphasizes. Freezing attention removes it. At second order, the corresponding decomposition is

$$\partial_r^2 a_i = \sum_j \left[\alpha_{ij} \partial_r^2 v_j + 2(\partial_r \alpha_{ij})(\partial_r v_j) + (\partial_r^2 \alpha_{ij}) v_j \right]. \tag{37}$$

These identities make temporal gluing part of the differential structure rather than a preliminary averaging step.

The learned frequencies provide temporal scales available to the model. They do not force the learned kernel to depend only on time differences. Calendar time or another chosen origin can remain relevant. In particular, a nonzero linear component is injective as a scalar time feature, although a subsequent projection can discard it. Finite periodic components alone can have aliases. The scientific meaning of time transfer therefore includes its units and origin as well as the learned parameters.

7.3 Restriction and the defect of local assembly

Attention also reveals why gluing across observation contexts is a substantive problem. Let R_U retain the tokens in a subrecord U. In general,

$$R_U A_\theta(h) V_\theta(h) \neq A_\theta(h|_U) V_\theta(h|_U). \tag{38}$$

Even in the first layer, where retained values are unchanged, removing tokens changes the softmax denominator. If omitted tokens carry total attention weight η and the full-history head has retained and omitted weighted means $\bar{v}_U, \bar{v}_O$, then

$$a_i^{\text{full}} = (1-\eta)\bar{v}_U + \eta \bar{v}_O, \qquad a_i^{\text{restricted}} - a_i^{\text{full}} = \eta(\bar{v}_U - \bar{v}_O). \tag{39}$$

Later layers propagate this difference through contextual values and scores. Missing event history in a system of records can therefore change the representation even when all retained codes match.

We use "gluing" for this explicit assembly and reserve exact descent for assemblies whose induced local realizations satisfy compatibility on overlaps. A stochastic attention matrix need not be invertible and need not satisfy a cocycle identity. The categorical condition concerns the represented observations and operators after assembly. Theorem 5.1 states exact compatibility at the profile level; Section 10 adds the preservation of terminal dynamics.

8 Tropical local structure and differential interfaces

A finite feedforward ReLU map is continuous and piecewise affine. Fixing an activation pattern σ gives

$$F(z) = A_\sigma z + b_\sigma, \qquad z \in C_\sigma, \tag{40}$$

where the cells form a finite polyhedral subdivision. A scalar output admits a difference-of-convex representation

$$f(z) = \max_i (a_i^\top z + c_i) - \max_j (b_j^\top z + d_j). \tag{41}$$

With real slopes this is the generalized tropical-rational description appropriate to real network weights. The integral-slope version belongs to the usual tropical algebra. Signed output weights make the difference in (41) necessary in general [12, 13].

The local structure applies to the inputs of a ReLU block or head. Softmax, sinusoidal time features and normalization mean that the complete temporal transformer is generally piecewise smooth rather than globally piecewise affine. On a fixed-length input stratum with fixed token types and mask, regular activation interfaces divide the continuous variables into smooth regions. A feature-space hyperplane can pull back through attention to a curved interface. Both the smooth interior derivatives and the interface contributions must be retained.

Proposition 8.1 (Distributional interface contribution)**.** *Let f be continuous and piecewise C^2 on a domain, with a locally finite collection of smooth interfaces, one-sided gradient traces, and locally integrable piecewise second derivatives. At a regular interface S, choose a unit normal n pointing from the minus to the plus side. Distributionally,*

$$D^2 f = (D^2 f)_{\mathrm{reg}}\,\mathrm{d}x + \sum_S [\nabla f]_S \otimes n_S\,\mathcal{H}^{d-1}\restriction_S, \tag{42}$$

provided the interface terms have locally finite total variation. For a finite piecewise-affine map the regular part is zero.

Proof. Continuity of f makes its first distributional derivative equal to its piecewise gradient without a surface delta. Apply integration by parts to that gradient on each region. The boundary terms from adjacent regions combine to the jump $[\nabla f]_S$ multiplied by the oriented normal. Continuity along S forces agreement of tangential derivatives, so the jump is normal and the displayed tensor is symmetric. Lower-dimensional junctions contribute no additional term under the stated regularity assumptions. □

For $f(z) = \mathrm{ReLU}(a^\top z + b)$ with $a \neq 0$, the formula becomes

$$D^2 f = aa^\top \delta(a^\top z + b) = \frac{aa^\top}{\|a\|}\,\mathcal{H}^{d-1}\restriction_{\{a^\top z+b=0\}}\,. \tag{43}$$

The Hessian vanishes at almost every ordinary evaluation point but is nonzero as a distribution. For a signed sum of hinges, its wall weights can have either sign. A sigmoid applied to a piecewise-affine logit additionally produces ordinary within-cell curvature and rescales the gradient jumps. Thus the differentiated scalar, whether logit, probability or residual time-to-termination, must be specified.

8.1 Finite perturbations retain the singular contribution

The useful quantity for a trajectory comparison is often a finite contrast rather than an infinitesimal derivative. For a scalar readout f along an admissible path $x(r)$, set $v(r) = f(x(r))$ and

$$\Delta_h v = v(h) - 2v(0) + v(-h), \qquad h > 0. \tag{44}$$

A path can be a straight latent displacement, a specified time perturbation or a realized event-edit family. Its definition is part of the experiment.

Theorem 8.2 (Exact contrast identity)**.** *If v is continuous and its distributional second derivative is a finite signed measure μ on a neighbourhood of $[-h, h]$, then*

$$\Delta_h v = \int_{-h}^{h} (h - |r|)\,\mu(\mathrm{d}r). \tag{45}$$

For a piecewise-affine restriction with slope jumps c_k at r_k,

$$\Delta_h v = \sum_{|r_k|<h} c_k (h - |r_k|). \tag{46}$$

Proof. The triangular kernel $k_h(r) = (h - |r|)_+$ has distributional second derivative $\delta_{-h} - 2\delta_0 + \delta_h$. Integration by parts twice, or approximation of this kernel by smooth compactly supported functions, gives $\langle v'', k_h\rangle = v(-h) - 2v(0) + v(h)$. The atomic form follows by substitution. □

Equation (45) retains smooth and singular contributions in the same expression. For a locally affine head, a finite perturbation detects every crossed wall with its precise weight and distance from the center. Automatic differentiation of the pointwise second derivative at an interior point need not detect any of them. For the full attention network, the measure contains the smooth contribution from (37) as well as activation-interface terms.

This identity also gives an error scale. If $|\widehat{v} - v| \le \varepsilon$ at the three evaluation points, then

$$|\Delta_h \widehat{v} - \Delta_h v| \le 4\varepsilon, \qquad \left|\frac{\Delta_h \widehat{v} - \Delta_h v}{h^2}\right| \le \frac{4\varepsilon}{h^2}. \tag{47}$$

Very small perturbations amplify prediction error. A claim about resolved curvature must therefore specify both the perturbation scale and the accuracy of the readout at that scale.

8.2 What survives stochastic averaging

The tropical representation class and max-plus linearity are different properties. Let $f_1, \ldots, f_m$ be finite continuous piecewise-affine functions on a common finite-dimensional domain and let $p_j \ge 0$ be constants summing to one. Then $\sum_j p_j f_j$ is piecewise affine on the common refinement of their subdivisions. The same holds for a finite stochastic transition with affine branches F_j and a piecewise-affine continuation value: $x \mapsto \sum_j p_j v(F_j x)$ remains in that class. A maximum over finitely many such action values remains piecewise affine as well.

The proof is constructive. Pull back each branch subdivision through its affine map, take their common refinement, and sum the affine expressions on each cell. To take a maximum, subdivide further along equality hyperplanes. Thus finite stochastic branching with fixed probabilities preserves a tropical-rational representation even though expectation does not preserve maxima.

Indeed, for a fair random sign $Y \in \{-1, 1\}$, $f(Y) = Y$ and $g(Y) = -Y$ give

$$\mathbb{E}[\max\{f, g\}] = 1 > 0 = \max\{\mathbb{E}f, \mathbb{E}g\}.$$

For continuous noise, even piecewise-affine closure can fail. If Z is uniform on $[-1, 1]$, then

$$\mathbb{E}[(x + Z)_+] = \frac{(x+1)^2}{4}, \qquad -1 < x < 1. \tag{48}$$

The averaged hinge has ordinary nonzero curvature in the interior. State-dependent mixture weights add further derivative terms; attention supplies the explicit example (37).

A finite-scenario approximation recovers piecewise-affine closure under the stated affine-branch assumptions, but its error must be controlled. If v is L-Lipschitz and the true and approximating transition laws have Wasserstein distance at most η uniformly over the states being compared, their expected values differ by at most $L\eta$, directly from the coupling definition. Near a singular boundary there may be no uniform L; finite-scale value bounds such as (47) are then the appropriate requirement. The unresolved problem is to establish approximation control for the actual stochastic operator at the boundary, not whether expectation is max-plus linear.

9 Absorption is part of the continuation operator

Activation interfaces and a terminal boundary play different mathematical roles. The former belong to a representation; the latter terminates the process being represented. Their relationship is established when a learned readout resolves the value and response structure imposed by the terminal condition.

Let X_t be a Markov realization on an alive domain Ω, with exit time ζ and an absorbing state $\dagger$. Its killed semigroup is

$$(P_t^\dagger f)(x) = \mathbb{E}_x[f(X_t)\mathbf{1}_{\{t<\zeta\}}]. \tag{49}$$

It is sub-Markov: $P_t^\dagger 1 \le 1$, with

$$S_t(x) = P_t^\dagger 1(x) = \mathbb{P}_x(\zeta > t), \qquad V(x) = \mathbb{E}_x\zeta = \int_0^\infty P_t^\dagger 1(x)\,\mathrm{d}t. \tag{50}$$

The last identity holds as a nonnegative extended integral; finite expected life is an additional integrability condition. A finite-horizon counterpart uses $\int_0^T S_t\,\mathrm{d}t$. Under the usual generator-domain and integrability assumptions the residual-life value solves $\mathcal{L}V = -1$ with the absorbing condition.

For a jump process, absorption includes overshoots. Consider a symmetric jump density $J(x, y)$ and zero extension $\widetilde{f}$ outside Ω. Whenever the principal-value expression is defined,

$$\begin{aligned} \mathcal{L}^\dagger f(x) &= \mathrm{PV}\int_{\mathbb{R}^d}(\widetilde{f}(y) - f(x))J(x, y)\,\mathrm{d}y \\ &= \mathrm{PV}\int_\Omega (f(y) - f(x))J(x, y)\,\mathrm{d}y - q_\Omega(x)f(x), \\ q_\Omega(x) &= \int_{\Omega^c} J(x, y)\,\mathrm{d}y. \end{aligned} \tag{51}$$

The exterior contributes a precise killing term at every interior point. Removing the exterior and retaining only the regional integral changes the operator. Conditioning transitions on survival changes it again. Neither operation is an innocuous restriction of the same mortality problem.

For $J(x, y) = c_{d,\alpha}\|x - y\|^{-d-\alpha}$, the boundary contribution has an explicit scale. In the half-space $\Omega = \{x_d > 0\}$ and at $x = se_d$, substitution $y = sz$ gives

$$q_\Omega(se_d) = c_{d,\alpha}s^{-\alpha}\int_{z_d\le 0}\|e_d - z\|^{-d-\alpha}\,\mathrm{d}z = C_{d,\alpha}s^{-\alpha}. \tag{52}$$

The integral is finite and positive for $\alpha > 0$. Thus the effect of possible exit is present before the process reaches the geometric boundary. This is one direct sense in which boundary treatment changes the interior dynamics.

9.1 Boundary asymptotics and the chosen generator

For the killed fractional Laplacian of order $\alpha \in (0, 2)$, regularity results on suitable domains identify the boundary factor $s^{\alpha/2}$, where s is distance to the boundary [14]. To state a differentiated asymptotic, write $\beta = \alpha/2$ and assume along an inward normal that

$$V(s) = cs^\beta + R(s), \quad c > 0, \qquad R^{(j)}(s) = o(s^{\beta-j}), \quad j = 0, 1, 2. \tag{53}$$

Then

$$V'(s) \sim c\beta s^{\beta-1}, \qquad V''(s) \sim c\beta(\beta - 1)s^{\beta-2}. \tag{54}$$

The sign is negative for $0 < \beta < 1$. The index-one case gives $V \sim c\sqrt{s}$ and $V'' \sim -(c/4)s^{-3/2}$. This is the boundary regime used in the lifecycle programme; the general exponent is $\alpha/2 - 2$, not a universal exponent for all absorbing processes.

The derivative bounds in (53) carry real content. A leading asymptotic for V alone does not justify differentiating its remainder. Nor does a finite event vocabulary by itself determine a stable jump law. For mixed drift, diffusion and jump generators, the boundary regime depends on their relative orders and degeneracy rates. The killed stable model makes the mechanism explicit and supplies a reference class against which an effective generator can be examined.

For a regular change of boundary coordinate $s = \varphi(\widetilde{s})$ with $\varphi(0) = 0$, $\varphi'(0) = a > 0$ and bounded φ'',

$$V_{\widetilde{s}\widetilde{s}} = V_{ss}(\varphi')^2 + V_s\varphi'' \sim c\beta(\beta - 1)a^\beta \widetilde{s}^{\beta-2}. \tag{55}$$

The exponent survives this class of coordinate changes. Arbitrary nonlinear latent reparameterizations need not preserve Euclidean Hessians or boundary distance. Finite responses along transported paths provide the more direct invariant.

9.2 Degenerate coordinates must transport the operator

A boundary exponent by itself is not invariant under a coordinate map that degenerates at the boundary. For the explicit change $y = s^m$, $m > 0$, the profile $V(s) = cs^\beta$ becomes $\widetilde{V}(y) = cy^{\beta/m}$. Its exponent has changed, while every value at corresponding states remains the same.

The operator changes with it. For a one-dimensional alive domain $(0, \infty)$, set $s = y^{1/m}$ and $z = \eta^{1/m}$ in the interior part of the pure-jump operator. The transformed interior integral has kernel

$$\widetilde{J}(y, \eta) = J(y^{1/m}, \eta^{1/m})\frac{1}{m}\eta^{1/m-1}, \tag{56}$$

with the principal-value prescription transported from the original integral when required. Any compensation convention must be transported as well. The transformed killing rate is $\widetilde{q}(y) = q(y^{1/m})$. A translation-invariant stable kernel generally becomes an inhomogeneous kernel. Keeping the original stable operator while changing only the value exponent would describe a different problem.

Thus localization can preserve the boundary-value problem under a degenerate coordinate map without preserving a Euclidean distance exponent. The invariant is the transported operator, its admissible function domain, its terminal data and the values along corresponding perturbation paths. Fractional boundary regularity concerns singular behavior of a solution on a specified domain; it does not require a degenerate coordinate transformation.

9.3 A boundary profile resolved by ReLU hinges

Consider the explicit profile

$$V_\beta(s) = s^\beta, \qquad s \ge 0, \quad 0 < \beta < 1, \tag{57}$$

with $V_\beta(0) = 0$. It isolates the normal boundary factor within a multidimensional absorbing process. For $0 < h < s$, define the improvement $G = V_\beta(s+h) - V_\beta(s)$ and the deterioration loss $L = V_\beta(s) - V_\beta(s-h)$. Concavity gives

$$L - G = 2s^\beta - (s-h)^\beta - (s+h)^\beta > 0. \tag{58}$$

For $h \ll s$,

$$L - G = \beta(1-\beta)s^{\beta-2}h^2 + O(h^4 s^{\beta-4}). \tag{59}$$

When $h = rs$, $0 < r < 1$, the exact expression is

$$L - G = s^{\beta}\big[2 - (1-r)^{\beta} - (1+r)^{\beta}\big]. \tag{60}$$

Consequently the normalized contrast diverges as $s^{\beta-2}$ while the absolute contrast tends to zero at fixed relative step r. The distinction specifies what boundary amplification means and avoids extrapolating a local quadratic expansion through the boundary.

Take knots $0 = s_0 < s_1 < \cdots < s_N$ and let

$$m_k = \frac{s_{k+1}^{\beta} - s_k^{\beta}}{s_{k+1} - s_k}, \qquad 0 \le k < N.$$

The linear interpolant is exactly a one-hidden-layer ReLU function:

$$V_N(s) = m_0 s + \sum_{k=1}^{N-1} (m_k - m_{k-1})(s - s_k)_+, \qquad 0 \le s \le s_N. \tag{61}$$

Its distributional second derivative on $(0, s_N)$ is

$$V_N'' = \sum_{k=1}^{N-1} (m_k - m_{k-1})\delta_{s_k}. \tag{62}$$

All coefficients are negative because the profile is concave. The boundary response is represented by weighted slope changes rather than by a smooth pointwise Hessian.

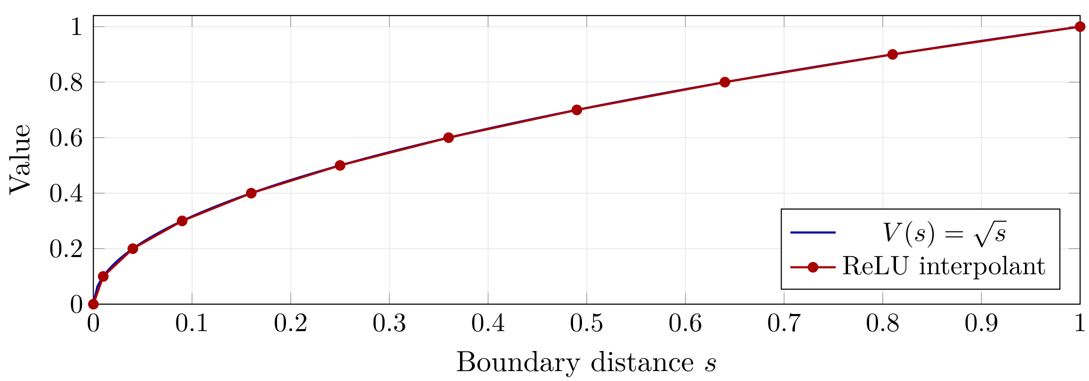

Figure 2: Analytic construction. The interpolant represents the profile by a finite collection of slope changes. Its classical Hessian is zero between knots; its finite contrasts retain the weighted atoms in (62).

On a fixed interval $[\varepsilon, 1]$, mesh refinement gives uniform convergence to s^{β} and distributional convergence of the second derivatives. At $s = 0$, the unbounded slope prevents a fixed finite interpolant from resolving arbitrarily small scales. A sequence of refined networks can approximate the profile at progressively finer resolution. No infinite wall density is attributed to a single finite network.

The usual linear-interpolation error on a cell $[a, b] \subset (0, 1]$ obeys

$$\|V_{\beta} - V_N\|_{\infty,[a,b]} \le \frac{(b-a)^2}{8}\,\beta(1-\beta)a^{\beta-2}. \tag{63}$$

Thus an absolute error target suggests cell widths of order $a^{1-\beta/2}$ in this bound, while relative resolution of the boundary profile calls for widths proportional to a. The first cell touching zero requires separate treatment. This calculation gives a reason for local refinement near the boundary without asserting that training automatically discovers the required mesh.

The construction links the boundary-value description to the tropical one. A fractional boundary profile can be evaluated through exact finite contrasts of a ReLU approximation. Temporal attention determines which contextual feature state is supplied to that approximation and how the state changes when event timing changes. The same contrast must include that contextual change whenever the experiment perturbs the history rather than only the final latent coordinate.

10 Localization between observational realizations

Let R denote a category of represented systems. An object contains a state realization, relevant readouts, continuation operators and terminal structure. An observation source U has an observation map and an encoder producing a local realization X_U. Source localization is the realization of the shared object in that observational setting. Recovering the common object from local data is the inverse problem.

On a domain of common represented states, suppose the realizations admit an invertible comparison $J_{UV} : X_U \to X_V$. For a scalar readout f_V define $J_{UV}^* f_V = f_V \circ J_{UV}$. Exact operator compatibility is

$$P_t^U J_{UV}^* = J_{UV}^* P_t^V. \tag{64}$$

With terminal structure retained, the same equation must hold for killed operators. Preservation of an ordinary transition mean is weaker than (64); future values can be nonlinear, and survival depends on missing transition mass.

Theorem 10.1 (Boundary-preserving localization)**.** *Suppose $J : X_U \to X_V$ is a bimeasurable bijection and*

$$P_t^{U,\dagger} J^* = J^* P_t^{V,\dagger} \qquad (t \geq 0) \tag{65}$$

on bounded measurable functions. Then the survival probabilities agree under J. Expected residual-life values agree as extended nonnegative functions, and agree as finite values wherever the integrals are finite. Discounted resolvents satisfy the same intertwining. If $x_U(r)$ is an admissible path and $x_V(r) = J(x_U(r))$, finite contrasts of corresponding readouts are identical along these paths.

Proof. Apply (65) to the constant function 1. This proves survival preservation. Integrate the resulting equality over t using Tonelli's theorem to obtain residual-life preservation. Multiplication by $e^{-\rho t}$ and integration gives the resolvent relation for bounded functions and $\rho > 0$. Readout preservation at the three path points proves equality of their finite contrasts. □

This theorem identifies an exact transfer target: the killed represented process, together with its observables. It is stronger than transferring a risk ranking. It also avoids treating different patient populations as matched individuals. The map compares the states and operators they realize on common support. Public data cannot recover distinctions that its observation channel removes unless other information supplies them.

The following elementary counterexample shows why the boundary condition belongs in the theorem. A system with one alive state and death rate $q > 0$ has

$$P_t^\dagger f = e^{-qt} f, \qquad V = \frac{1}{q}. \tag{66}$$

Conditioning on survival makes the transition operator the identity for every q. All these systems then appear identical, although their expected lifetimes differ. Exact knowledge of the survivor-conditioned operator cannot recover the missing killing rate.

Attention normalization is an operation on representation weights, not a physical survival law. Nevertheless the example explains why a row-normalized representation operator cannot be substituted for a killed transition operator without an additional survival readout. If attention is used to parameterize transition destinations, their survival mass must be represented separately or through an explicit terminal destination.

10.1 Approximate intertwining and value error

Let P, Q be discrete-time killed kernels and J^* pullback by a bimeasurable state bijection. Suppose, in supremum operator norm,

$$\|PJ^* - J^*Q\| \leq \varepsilon. \tag{67}$$

Since both kernels are contractions,

$$P^n J^* - J^* Q^n = \sum_{k=0}^{n-1} P^{n-1-k}(PJ^* - J^*Q)Q^k$$

has norm at most $n\varepsilon$. Hence n-step survival differs by at most $n\varepsilon$. For discounted residual continuation $V_P = \sum_{n\geq 0} \delta^n P^n 1$ and $0 < \delta < 1$,

$$\|V_P - J^*V_Q\|_\infty \leq \frac{\delta\varepsilon}{(1-\delta)^2}. \tag{68}$$

For an N-term undiscounted value the corresponding bound is $\varepsilon N(N-1)/2$. Infinite-horizon undiscounted comparison needs additional tail control. These estimates follow by summing the telescoping bound, so their dependence on the horizon is explicit.

Together with Theorem 8.2, this gives an operator-to-boundary-response estimate. If V_P, V_Q are evaluated along transported three-point paths,

$$\left| \frac{\Delta_h V_P - \Delta_h(V_Q \circ J)}{h^2} \right| \leq \frac{4\delta\varepsilon}{h^2(1-\delta)^2}. \tag{69}$$

Operator agreement at a fixed tolerance does not resolve arbitrarily small perturbations or an arbitrarily long value horizon. This is a second reconstruction bound, complementary to the Gromov–Hausdorff estimate: the first controls decoded state geometry, while this one controls a particular family of consequences of continuation.

Equation (67) is a strong uniform assumption. Empirical comparison on a finite set of readouts provides a restricted defect, not automatically the displayed norm. It becomes a certificate for iterated valuation when the tested function family is closed under the required backward operations or when a separate approximation bound controls the missing functions. This is the preservation requirement for transferring a representation into a new valuation problem.

11 Valuation under a sufficient reduction

Once a representation supports continuation, forward evolution and backward evaluation become two aspects of the same structure. A kernel $K : X \rightsquigarrow Y$ sends distributions forward by $\mu \mapsto \mu K$ and functions backward by

$$(K^*f)(x) = \int_Y f(y)K(x, \mathrm{d}y).$$

For $L : Y \rightsquigarrow Z$,

$$(L \circ K)^* = K^* L^*, \qquad \int_Y f \, \mathrm{d}(\mu K) = \int_X K^* f \, \mathrm{d}\mu. \tag{70}$$

The reversal of composition is exact. It does not require invertible dynamics. Markov categories provide a general categorical language for these probabilistic operations and notions of sufficiency [15].

A measurable reduction $q : H \to X$ intertwines kernels when

$$q_* K_a(h, \cdot) = \overline{K}_a(q(h), \cdot). \tag{71}$$

The profile construction gives this relation for the controlled history process of Section 3. For a learned state proposed independently, it is a condition to establish. Equality of next-state means is insufficient when future values are nonlinear.

Suppose a finite-horizon control problem has common finite feasible actions, bounded measurable rewards, and terminal payoff. Assume the rewards and terminal payoff factor through q and (71) holds at each stage. Write $q^* v = v \circ q$ and

$$(T_k v)(h) = \max_a \left\{ r_k(h, a) + \beta_k \int v(h') K_{k,a}(h, \mathrm{d}h') \right\}.$$

Proposition 11.1 (Valuation commutes with exact reduction)**.** *Under these assumptions, $T_k q^* = q^* \overline{T}_k$. Full-history optimal values equal the pullbacks of reduced optimal values at every stage.*

Proof. Substitute $v(q(h'))$ into the full integral, use (71), and replace each reward by its reduced readout. The same finite action maximum is then taken on both sides. Terminal values agree by assumption, and backward induction proves the assertion. □

The proposition explains why a representation can become scientifically useful beyond its original prediction target. If it supports the relevant continuation laws, additional value functions can be computed on the same state. Their values need not agree with a specific failure score, a failure probability, or the objective used to train the encoder. Reusability is a property of the preserved system, not of the numerical dimension of its outputs.

For a differentiable deterministic transition F, the local version of (70) is

$$\mathrm{d}(V \circ F)_x = (DF_x)^* \mathrm{d}V_{F(x)}.$$

With fixed actions and stage rewards, the co-state recursion is

$$p_k = \mathrm{d}r_k + \beta_k (DF_k)^* p_{k+1}, \qquad p_N = \mathrm{d}g. \tag{72}$$

The co-state evaluates the consequence of a state displacement for a specified objective. It is a covector, and its transport is the derivative of backward function transport. Compositional accounts of backpropagation and learners formalize closely related forward–backward structures [16, 17]. In a trained model, the loss adjoint and a control co-state coincide only when the relevant scalar objectives and computations coincide.

11.1 The condition that a lifecycle application must establish

The valuation proposition is a sufficiency statement about a specified joint process. A mortality state need not encode wealth, preferences, income or the response of health to spending. If those quantities affect future rewards and transitions, the reduced economic state must retain them. One possible state is $(a, q(h), z)$, where a is wealth and z contains the other required economic variables. The joint kernel, rather than its mortality marginal alone, must satisfy (71).

In the example of Section 2, the constant baseline score fails this requirement for the maintenance problem, while θ preserves its response probabilities and hence its value. This verifies the proposition in an explicit model and exhibits exactly how a reduction can fail. In clinical-lifecycle data the corresponding factorization and kernel conditions are empirical and model-specific. The proposition states the test; it does not establish its outcome.

11.2 Aggregation of exact finite responses

For individual paths $v_i(r)$ and scales $h_i > 0$, define $\kappa_i(h_i) = \Delta_{h_i} v_i / h_i^2$. Whenever the displayed moments exist,

$$\mathbb{E}[\Delta_{h_i} v_i] = \mathbb{E}[h_i^2]\mathbb{E}[\kappa_i(h_i)] + \mathrm{Cov}(h_i^2, \kappa_i(h_i)). \tag{73}$$

This identity is exact at the chosen scales, retaining both ordinary and interface contributions. It requires no Taylor remainder. The average of paired gains is half the central contrast; the worsening-minus-improvement convention reverses its sign. A population-average mortality rate does not determine this covariance.

If a boundary-distance density behaves as $p(s) \asymp s^{\eta}$ and $|V''(s)| \asymp s^{\beta-2}$, then the local absolute moment is governed by $\int_0^{\varepsilon} s^{\eta+\beta-2}\,\mathrm{d}s$ and is finite precisely when $\eta + \beta > 1$. Exposure scaling can change integrability of the weighted moment. Eventual absorption alone does not determine the aggregate size of the boundary effect.

12 Modules, differential gluing and the localization theorem

The relation to Beilinson–Bernstein localization becomes precise when a represented system carries an algebra of operations. An embedding vector by itself is not a module: the module is a space of represented observables or sections together with an action of that algebra. Embedding coordinates and decoded responses can generate such a space.

Let R be an algebra generated by selected continuation and readout operations, with their relations, acting on a vector space M. This definition requires closure under those operations; the raw finite embedding coordinates need not be closed under arbitrary continuation. A finite-dimensional invariant space is a stronger property than a finite-dimensional nonlinear encoding.

Suppose a geometric realization X carries a sheaf of algebras $\mathscr{A}$ with a compatible map from R to its global sections. There is a canonical localization construction

$$\mathrm{Loc}_X(M) = \mathscr{A} \otimes_R M, \tag{74}$$

where the tensor product is sheafified and uses the induced right R-action. For a left $\mathscr{A}$-module $\mathcal{M}$,

$$\mathrm{Hom}_{\mathscr{A}}(\mathrm{Loc}_X M, \mathcal{M}) \simeq \mathrm{Hom}_R(M, \Gamma(X, \mathcal{M})). \tag{75}$$

An R-linear map into global sections extends by $a \otimes m \mapsto a \cdot s_m$; restriction to $1 \otimes m$ gives the inverse. The unit and counit are

$$M \longrightarrow \Gamma(X, \mathrm{Loc}_X M), \qquad \mathscr{A} \otimes_R \Gamma(X, \mathcal{M}) \longrightarrow \mathcal{M}. \tag{76}$$

Localization is right exact in general. The unit and counit being isomorphisms is the substantive recovery condition.

In Beilinson–Bernstein theory, X is a flag variety, R is the relevant central reduction of the universal enveloping algebra, and $\mathscr{A}$ is a sheaf of twisted differential operators. Suitable regularity and dominance conditions turn localization and global sections into inverse equivalences [18, 19]. Equation (75) supplies the same formal architecture for a represented system once its algebras are specified. The flag-variety theorem does not automatically establish the recovery conditions for an observational dataset.

For realizations obtained through different observation channels, the proposed reconstruction is therefore an algebraic task: identify the shared operational representation, realize its action locally, and determine the subcategory on which (76) is invertible. When this succeeds, the cross-source passage is

$$\mathcal{M}_U \longmapsto \Gamma(X_U, \mathcal{M}_U) \longmapsto \mathrm{Loc}_{X_V}\, \Gamma(X_U, \mathcal{M}_U). \tag{77}$$

This is a precise localization construction rather than a direct replacement of one vocabulary by another.

12.1 Differential modules retain the walls

On a smooth feature domain, let $\mathcal{D}_X$ be the sheaf of finite-order differential operators with smooth coefficients. A locally integrable ReLU readout defines a distribution f. The cyclic module

$$\mathcal{D}_X f = \{Pf : P \in \mathcal{D}_X\} \tag{78}$$

retains its derivatives, including the interface distributions in (42). Multiplication by smooth coefficients is well-defined on distributions. Taking only ordinary derivatives inside open activation cells discards part of this module's action.

The smooth differential setting here is a working realization. Algebraic or analytic $\mathcal{D}$-module theory imposes its own coefficient category and finiteness conditions, including those required for holonomicity. A finite ReLU expression alone does not settle those conditions. A nonlocal killed generator, such as (51), is also not a finite-order differential operator. The full operational algebra must contain it separately, through integral operators or semigroup operators. Its local differential part still controls sensitivities and interface contributions.

12.2 What a twist records

Suppose local scalar frames are related by nowhere-vanishing functions g_{UV} satisfying the multiplicative cocycle condition. The local differential operators glue by conjugation:

$$P_V = g_{UV}^{-1} P_U g_{UV}, \qquad \partial_r \longmapsto \partial_r + g_{UV}^{-1} \partial_r g_{UV}. \tag{79}$$

This constructs the familiar line-bundle example of a twisted differential-operator sheaf. More general TDOs extend this construction. Matrix-valued changes of frames give gauge transformations of a vector bundle and must be distinguished from a scalar TDO twist [19].

For a local invertible matrix frame $G(x)$,

$$G^{-1}\partial_r G = \partial_r + \mathcal{A}_r, \qquad \mathcal{A}_r = G^{-1}(\partial_r G), \tag{80}$$

where the left side denotes conjugation as an operator. It follows that

$$G^{-1}\partial_r^2 G = \partial_r^2 + 2\mathcal{A}_r\partial_r + \partial_r\mathcal{A}_r + \mathcal{A}_r^2. \tag{81}$$

Thus an exact change of representation can create first- and zeroth-order correction terms while preserving the represented operator. A pure change of frame gives a locally flat connection; nonzero curvature requires additional connection data, not merely a change of coordinates.

Time2Vec-conditioned attention provides explicit, differentiable coefficients from which local comparisons may be constructed. Its derivatives are given by (34)–(37). To interpret an induced comparison as (79) or (80), one needs an invertible frame map on the relevant domain and operator compatibility on overlaps. The attention weight matrix itself need not provide either. This formulation preserves the proposed role of attention in gluing while stating the exact additional conditions for a differential twist.

The boundary makes those conditions consequential. A transported operator must act correctly on the singular part as well as the ordinary interior part. If it also contains killing, its transported action must preserve the lost mass. The resulting object consists of local differential representations, their temporal compatibility, and the terminal operator that fixes continuation.

13 Learning, coverage and computable separating experiments

The response law defining the canonical profile is unknown in applications. Learning estimates it on an observed history–test distribution, and experiment selection determines which additional distinctions can be resolved. These tasks are separate from the universal-property theorem.

For binary test outcome Y_t, let $p_h(t) = \mathbb{E}[Y_t \mid h, t]$ under the specified observation or assignment law. The population Brier risk satisfies

$$\mathcal{L}(f) - \mathcal{L}(p) = \mathbb{E}_{h,t}(f(h,t) - p_h(t))^2. \tag{82}$$

The cross term vanishes by conditional expectation. This is an average error over supported pairs, whereas Theorem 4.4 assumes a uniform error on its comparison set. An error of one on a set of mass η contributes only η to this risk. Subgroup and boundary coverage therefore matter even when average prediction is accurate.

13.1 A finite model-discrimination criterion

A computable design can be specified using candidate models $M_1, \ldots, M_K$, positive prior weights π_k, and their joint likelihoods $L_k(D)$ for the available data. Within this finite model class, define

$$w_k(D) = \frac{\pi_k L_k(D)}{\sum_j \pi_j L_j(D)}.$$

Joint likelihoods must respect temporal dependence; a product of unrelated marginal probabilities is not a substitute. Let a finite candidate experiment set $\mathcal{E}_0$ have finite outcomes y, with candidate predictions $p_k(y \mid h, e)$, and write $\bar{p} = \sum_k w_k p_k$. The expected information gain about the model index is

$$I(e \mid h, D) = \sum_{k,y} w_k p_k(y \mid h, e) \log \frac{p_k(y \mid h, e)}{\bar{p}(y \mid h, e)}. \tag{83}$$

Using the usual zero-mass convention, this finite sum is directly computable in $O(K|\mathcal{Y}|)$ operations per experiment once predictions are available. One can maximize $I(e \mid h, D) - \lambda c(e)$ over feasible candidates, where $c(e)$ is cost. This is the finite-model version of Bayesian experimental design, not a tractable solution to an unrestricted supremum over all laws [20].

For the two-mode example, baseline success has the same probability in both modes and therefore gives zero information about the mode. The diagnostic experiment has positive information whenever both posterior mode weights are positive. Thus a concrete design calculation chooses a test that refines the collapsed baseline state. The conclusion is conditional on the candidate model class: models can agree because they all omit the relevant mechanism.

More generally, enlarging an experimental family $\mathcal{E} \subseteq \mathcal{E}'$ gives a canonical surjection $Q_{\mathcal{E}'} \to Q_{\mathcal{E}}$. A family of current readouts may need closure under supported continuation before it defines a state with closed updates. Experiment selection can discover missing distinctions; it cannot certify untested continuation solely from agreement on the current family.

13.2 Which response law is identified

Prescribed actions and observed actions are different test specifications. Two models can have the same observational joint law and different intervention laws, so no categorical quotient of the former identifies the latter without additional information. Controlled tests already supported by empirical evidence can be included directly. The reconstruction theorem then applies to that enlarged response law.

Observation channels matter as well. If O_U is a source-specific recording channel applied to an underlying law P, the observed law is $(O_U)_*P$. Cross-source localization requires a common operational interpretation of the compared responses. Differences in recording frequency, time windows or terminal ascertainment may change this interpretation. The comparison should therefore report decoded response error, continuation defect and survival defect on their stated support, rather than treating coordinate alignment as a complete transfer certificate.

14 A common system across the research programme

The four manuscripts develop successive parts of a single construction. Discrete causal topology supplies histories and candidate state representations [2]. Financial event sequences expose temporal distinctions that conventional summaries erase [3]. The mortality boundary supplies termination and residual-life values [4]. Lifecycle valuation places those values inside a controlled joint process [5].

Their common object is a system of distinguishable continuations. This makes the connection stronger than a shared choice of encoder: the represented state must support the operations needed by the next problem. A financial summary that erases a separating continuation incurs the obstruction of Proposition 1.1. A mortality representation that discards killing loses residual life even when survivor-conditioned transitions agree. An economic reduction must retain the joint variables and action-dependent kernels needed for valuation, as in Proposition 11.1.

Heterogeneous histories make these requirements observable. The response to a specified progression depends on the state receiving it, and its aggregate consequence depends on the joint distribution of responses and exposures in (73). Different collection processes then supply local views with which to examine whether this structure survives a change of observation channel. Common decoded responses establish one level of agreement; continuation and boundary preservation establish the levels required for dynamics and valuation.

The empirical task is consequently organized around separating responses, local comparison defects and operator defects on stated support. The mathematical results specify how each defect propagates into geometry or value. This connects the experiments to the theory without requiring every dataset to expose every operation of the full system.

The same reconstruction problem can arise in other event-driven domains. What transfers is the requirement that a state preserve the operations needed to understand and act on its trajectories.

15 The methodological relation to causal-set reconstruction

Causal-set theory asks how geometry can be constrained by discrete order and number, including whether a suitable continuum approximation exists and whether it is essentially unique [21]. The learned-state problem asks which effective geometry is fixed by the continuation laws of observed histories. Both require a choice of admissible realizations and an observation map that distinguishes them.

Bombelli compares Lorentzian geometries through distributions $P_n(C \mid G)$ of finite causal orders obtained by sampling [22]. A representative comparison is

$$D_n(G, G') = \frac{2}{\pi} \arccos \left(\sum_C \sqrt{P_n(C \mid G) P_n(C \mid G')} \right).$$

Our metric (15) instead takes a discounted supremum of response differences over the countable family of finite tests. These are different metrics on different objects. Their common method is comparison through observation laws without an arbitrary prior matching of coordinates.

The operational completion theorem does not identify a Lorentzian geometry or prove the causal-set Hauptvermutung. A physical identification would additionally require that the relevant observation law separate the proposed physical realizations. The useful transfer of ideas is the insistence on reconstruction criteria, scale and invariance. For the present programme, boundary-preserving operator comparison supplies an additional criterion that can be evaluated independently of any spacetime interpretation.

16 Discussion

The reconstruction problem begins with operations on histories. A representation becomes a state when it supports the response laws and continuations by which the system is distinguished. The terminal profile realization identifies the common object, and its universal property fixes the comparisons between reduced descriptions. This gives categorical reconstruction a concrete role in deciding which changes of representation preserve a system.

Temporal direction, geometry and localization follow from that requirement. Response-relevant order cannot be discarded by an exact realization. The operational metric measures the distinctions retained at a chosen resolution, while the finite reconstruction bound separates estimation error from unresolved continuation. Compatible local realizations inherit their cocycle from uniqueness, so their agreement concerns the represented system as well as its coordinates.

Temporal attention and ReLU structure make these requirements computationally explicit. Learned time signatures influence which events are coupled and how their contextual contributions change. Activation interfaces carry finite response through slope changes. The architecture therefore supplies an object on which compatibility and response can be calculated, including the effects of observation changes that alter the available context.

At a terminal boundary, preservation must extend to lost probability mass. Intertwining the killed operators retains termination laws and the values derived from them. Finite contrasts provide a corresponding comparison of response along transported paths, including singular contributions that an interior Hessian can miss. A boundary is thus part of the reconstructed dynamics throughout the domain where it affects continuation.

The framework links these results through one demand: the operations that make a system intelligible must survive its representation. A reusable learned state is one on which further observations, continuations and values remain consistent. Reconstructing such a state makes geometry a consequence of the system's distinguishable behavior.

References

[1] V. Zhorin, *Tensor-Based Computing in Contract Theory, IO, and HJB PDE Macro Models*, presentation dated April 16, 2015. doi:10.13140/RG.2.2.35041.10084.

[2] V. Zhorin, *Discrete Causal Topology*, manuscript (2025). Full text.

[3] V. Zhorin, *Why Financial Event Sequences Require Temporal Transformers: Necessity Theorems for Default Prediction*, manuscript. Full text.

[4] V. Zhorin, *Mortality Prediction as Boundary Value Problem*, manuscript (2026). Full text.

[5] V. Zhorin, *The Mortality Input Problem: Trajectory-Dependent Death and the Lifecycle Model*, revised manuscript (September 2026).

[6] C. R. Shalizi and J. P. Crutchfield, *Computational mechanics: pattern and prediction, structure and simplicity*, Journal of Statistical Physics **104**, 817–879 (2001). arXiv:cond-mat/9907176.

[7] S. Singh, M. R. James and M. R. Rudary, *Predictive state representations: a new theory for modeling dynamical systems*, UAI (2004), 512–519. arXiv:1207.4167.

[8] P. Baldan, F. Bonchi, H. Kerstan, and B. König, *Coalgebraic behavioral metrics* (2017). arXiv:1712.07511.

[9] S. Abramsky and A. Brandenburger, *The sheaf-theoretic structure of non-locality and contextuality*, New Journal of Physics **13**, 113036 (2011). arXiv:1102.0264.

[10] S. M. Kazemi et al., *Time2Vec: Learning a Vector Representation of Time* (2019). arXiv:1907.05321.

[11] A. Vaswani et al., *Attention Is All You Need*, NeurIPS (2017). arXiv:1706.03762.

[12] L. Zhang, G. Naitzat and L.-H. Lim, *Tropical Geometry of Deep Neural Networks*, ICML (2018). arXiv:1805.07091.

[13] M.-C. Brandenburg, G. Loho and G. Montúfar, *The Real Tropical Geometry of Neural Networks* (2024). arXiv:2403.11871.

[14] X. Ros-Oton and J. Serra, *The Dirichlet problem for the fractional Laplacian: regularity up to the boundary*, Journal de Mathématiques Pures et Appliquées **101**, 275–302 (2014). arXiv:1207.5985.

[15] T. Fritz, *A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics*, Advances in Mathematics **370**, 107239 (2020). arXiv:1908.07021.

[16] B. Fong, D. I. Spivak, and R. Tuyéras, *Backprop as functor: a compositional perspective on supervised learning*, LICS (2019). arXiv:1711.10455.

[17] D. I. Spivak, *Learners' languages*, EPTCS **372**, 14–28 (2022). arXiv:2103.01189.

[18] A. Beilinson and J. Bernstein, *Localisation de $\mathfrak{g}$-modules*, Comptes Rendus de l'Académie des Sciences, Série I **292**, 15–18 (1981). Original publication.

[19] M. Kashiwara, *Representation theory and D-modules on flag varieties*, Astérisque **173–174**, 55–109 (1989). Full text.

[20] J. Beck, B. M. Dia, L. F. R. Espath, Q. Long and R. Tempone, *Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain* (2017). arXiv:1710.03500.

[21] S. Surya, *The causal set approach to quantum gravity*, Living Reviews in Relativity **22**, 5 (2019). doi:10.1007/s41114-019-0023-1.

[22] L. Bombelli, *Statistical Lorentzian geometry and the closeness of Lorentzian manifolds*, Journal of Mathematical Physics **41**, 6944–6958 (2000). arXiv:gr-qc/0002053.