

# When the Formalism Becomes the Interface

## A Self-Audit of Speculative Formal Systems

Flyxion

Independent Researcher

May 2026

### **Abstract**

Formal frameworks in speculative theoretical work face a distinctive failure mode: the internal elegance of a notation system can begin regulating admissible interpretation more strongly than contact with observation regulates the notation itself. When this occurs, the formalism has ceased to function as a tool for navigating reality and has become, instead, an interface through which reality is filtered and gradually displaced. The danger is not merely that a theory becomes incorrect, but that it becomes internally coherent in a way that obscures the loss of external constraint, allowing aesthetic continuity to masquerade as explanatory depth. Systems of notation, once sufficiently elaborate, can generate their own standards of intelligibility and begin stabilizing themselves against revision rather than remaining answerable to the phenomena they were built to describe.

This essay introduces an audit protocol for identifying when formal structure has crossed from necessary to aesthetic, examines several case studies drawn from my own framework development, and argues that a theory capable of recognizing this boundary in itself is in a more epistemically defensible position than one that cannot. The audit is not a demolition. It is a form of ongoing theoretical maintenance — the condition under which a living research program remains distinct from a finalized doctrine.

*When compressed representations cease being tools for navigating reality and instead become the reality that systems act upon, causal recoverability deteriorates.*

---

## **1. The Seduction of Beautiful Machinery**

There is a particular pleasure that arrives when a theoretical framework begins to cohere. Definitions click into place, analogies across domains resolve into structural identities, and the notation develops an internal momentum that feels like discovery. I have experienced this repeatedly across several years of developing what I call the Relativistic Scalar-Vector Plenum framework, a formal apparatus for describing systems in terms of admissibility fields, trajectory bundles, and constraint propagation. The experience of that coherence is not, in itself, evidence that the framework is correct or that it maintains honest contact with the phenomena it purports to describe.

This essay is about the gap between those two things: the internal coherence of a formal system and its contact with the world. The gap is not exotic. It appears in every speculative framework sophisticated enough that readers — and authors — stop routinely asking what each formal layer is actually doing. Once notation becomes sufficiently elaborate, its internal consistency becomes easier to verify than its explanatory necessity. The formalism begins to feel like the thing it was built to represent. And at that point, something has gone wrong in a way that is difficult to detect precisely because everything still appears to be working.

I want to name this failure mode carefully, because I believe I have reproduced it in my own work, and because the act of naming it inward — rather than prescribing its correction outward toward other frameworks — is the only intellectually honest version of the argument. An essay about theoretical self-examination written in the impersonal register would quietly reproduce the failure it diagnoses: a framework standing outside the system it analyzes, radiating neutral formal authority while remaining unexamined itself. The first person is not a stylistic choice here. It is an epistemic commitment. The framework is being placed under its own audit. That self-application is not incidental to the essay's argument. It is the central proof of seriousness.

## 2. Three Kinds of Structure

Before the audit can become operational, the taxonomy of failure must be precise. I distinguish three kinds of formal structure, defined by what happens when the structure is removed.

*Necessary structure* is load-bearing in the strict sense: removing it destroys the framework's ability to draw distinctions it claims to draw, generate constraints it claims to generate, or explain transitions it claims to explain. Without this structure, something the theory actually does becomes unavailable. The local-to-global coherence conditions in sheaf-theoretic constructions are necessary when the claim being made is genuinely about how local patches of information can be glued into a globally consistent picture. If that is the content of the claim, the formalism earns its place.

*Convenient structure* compresses derivations, clarifies organization, or makes certain arguments easier to communicate, but could in principle be replaced by simpler machinery without substantive loss. Much of the notational scaffolding in any working theoretical document belongs to this category. Convenient structure becomes problematic only when it is mistaken for necessary structure — when the ease of derivation it provides is interpreted as evidence of deep explanatory insight rather than notational efficiency.

*Aesthetic structure* creates a feeling of depth, rigor, or unification while contributing no indispensable constraint. It does not draw distinctions that could not be drawn without it, does not block derivations that would otherwise go through incorrectly, and does not make predictions that simpler machinery would fail to make. Aesthetic structure is not the same as beauty, and it is not necessarily useless. In early stages of framework development, elegant constructions can function heuristically: they guide exploration before their necessity is established, suggest directions that turn out to be productive, and provide provisional scaffolding while load-bearing elements are still being identified.

The pathology arises not from elegance itself but from what happens next. If heuristic elegance silently upgrades itself into ontological commitment — if the scaffolding becomes the building — then the framework has acquired a component that resembles necessary structure from the inside while performing no necessary function. The transition from heuristic to constitutive is rarely announced. It accumulates through repeated use, through the citation of aesthetic components in other papers as though they were established results, and through the gradual erosion of the original uncertainty that prompted the

construction. The result is machinery that is too beautiful to question.

Most formal frameworks slide continuously between these categories. This is not a moral indictment but an observation about how theories are built: exploratory, piece by piece, with retrospective organization that can obscure the original contingency of each component's introduction. The problem is not that this sliding occurs. The problem is that it is invisible unless actively audited.

### **3. Representational Sovereignty**

The specific failure mode I am diagnosing has a structure that appears not only in formal theoretical systems but across a much wider class of representational architectures. Understanding how it operates institutionally clarifies what is at stake when it operates theoretically.

Consider a key performance indicator dashboard designed to make an organization's operational state legible to its managers. The dashboard begins as a navigational tool: a compressed representation of a complex underlying reality, chosen because no one can act directly on the full causal complexity of an institution. The compression is useful precisely because it is lossy. It makes decision-relevant structure visible at the cost of discarding the rest.

The failure occurs when the organization begins optimizing primarily against the dashboard rather than the state it was built to track. Managers who control what the indicators measure can improve measured performance without improving actual performance. Departments whose work is legible to the dashboard grow; those whose work is not become invisible to resource allocation. Over time, the institution loses the vocabulary, the incentive structure, and the memory architecture required to detect deviation between what the dashboard shows and what is actually happening. The compressed surface has become operationally sovereign: it now regulates what counts as success, what counts as failure, and eventually what counts as real.

The same dynamic appears in social media engagement metrics, which compress the complex dynamics of human attention into a single optimization target, and in credit scores, which compress a person's financial history and social circumstances into a number that then regulates access to resources. In each case, the representation begins as an instrument of navigation and ends as an instrument of governance. The map does not merely obscure the territory. The territory becomes inaccessible, and the map generates its own standards of correctness. Goodhart's observation — that a measure ceases to be a good measure once it becomes a target — describes the symptom; the underlying

mechanism is that optimization pressure causes the institution to lose causal contact with the state the measure was built to track.

A formal theoretical framework can undergo an exactly analogous process. The notation begins as a navigational tool: a way of making structural relationships legible, of compressing derivations, of communicating claims precisely. As the framework develops, the notation acquires its own internal logic, its own aesthetic standards, its own criteria of correctness. These internal standards are useful because they are rigorous. But if they begin regulating admissible interpretations more strongly than contact with observation regulates the formalism itself, then the formalism has become sovereign over its subject matter rather than answerable to it.

I want to call this *formal representational sovereignty*: the condition in which a framework's internal criteria of coherence have displaced its external criteria of correctness. A formalism in this condition is not false. It may be internally consistent, derivationally productive, and aesthetically compelling. What it is not is epistemically alive in the sense that matters: it is no longer capable of being forced into structural revision by something outside itself.

#### **4. The Audit Protocol**

The audit I am proposing is not a generalized suspicion toward formal structure. It is an operational test with a specific criterion: formal structure is epistemically justified insofar as removing it measurably reduces the framework's ability to preserve distinctions, generate constraints, or explain transitions that simpler machinery cannot. That is the subtraction criterion, and it is deliberately severe.

The test proceeds in three steps. First: identify the formal component to be examined — a particular piece of mathematical machinery, a defined primitive, an ontological commitment encoded in the notation. Second: remove it, or replace it with the simplest alternative that handles the same apparent function. Third: ask whether any distinction the framework draws, any constraint it generates, or any transition it explains has become unavailable. If the answer is no across all three, the component is aesthetic, regardless of how much structural depth it appears to contribute from the inside.

The distinction between *parameter adjustment* and *structural revision* sharpens this further and is the audit's most diagnostically useful tool. A framework that can absorb arbitrary outcomes by adjusting parameters while preserving its deep architecture is not functioning as an explanatory theory in any strong sense. If no conceivable observation could force a change in the framework's

ontological commitments or its fundamental dynamic structure — if everything that happens can be accommodated by tweaking coefficients or adding auxiliary hypotheses — then the theory is operating more like an interpretive language than a constrained explanatory system. The theory can describe anything; it explains nothing in the sense of ruling out alternatives.

This is not identical to Popperian falsifiability, though the family resemblance is close. Popper's criterion targets a theory's empirical predictions. The subtraction criterion targets formal components within a theory — mathematical machinery, defined primitives, ontological commitments — and does not require that structural revision be forced by empirical observation specifically. Mathematical results, internal consistency failures, and successful derivations by simpler means all count as forcing conditions. But the underlying diagnostic logic is the same: a theoretical element that is immune to revision from any direction is not a finding. It is a commitment disguised as one.<sup>1</sup>

A procedural parallel from an entirely different domain helps clarify what the audit is doing. In his analysis of exploitation and domination, Vrousalis insists on asking, for each candidate theoretical account, what exact relation is doing the explanatory work: is it harm, unfairness, vulnerability, or something more specifically structural? His point is that different candidate formalisms often collapse distinct mechanisms into apparent unity, and that the collapse is visible only when you press the question of which relation is load-bearing.<sup>2</sup> That is exactly the question the subtraction criterion asks about formal components in speculative systems: which relation is doing the explanatory work, and what happens when it is removed?

Applied honestly, this audit is uncomfortable. I have applied it to several components of my own framework, and the results are uneven.

---

<sup>1</sup> The connection to Lakatos is also worth noting. Lakatos distinguished progressive from degenerative research programmes by asking whether a programme's theoretical core was generating novel predictions or merely protecting itself through auxiliary hypotheses. The subtraction criterion extends this question inward: it asks not only whether the programme is progressive relative to observation but whether each formal component within it is progressive relative to the distinctions and constraints the theory actually needs to draw. See Imre Lakatos, *The Methodology of Scientific Research Programmes* (Cambridge University Press, 1978).

<sup>2</sup> Nicholas Vrousalis, *Exploitation as Domination: What Makes Capitalism Unjust* (Oxford University Press, 2023). The methodological parallel is in Vrousalis's treatment of rival accounts of exploitation, not in his substantive political conclusions.

## 5. Case Studies in Load-Bearing and Decorative Formalism

### 5.1. *The Yarncrawler Sheaf Construction: Necessary Structure*

The Yarncrawler framework is a system for reconstructing world-states from incomplete, locally available information streams. Its central claim is a Sheaf-Variational Equivalence: that the problem of coherent global reconstruction from locally consistent patches is formally equivalent to a variational problem over a sheaf of information functors. The sheaf-theoretic machinery here is, I believe, load-bearing in the strict sense. The claim is not merely that local information can be aggregated; it is that local consistency conditions must satisfy specific gluing axioms for global reconstruction to be possible, and that failure of gluing corresponds to genuine obstructions in the reconstruction problem. Remove the sheaf structure and the distinction between “locally consistent but globally incoherent” and “genuinely globally coherent” collapses. The formalism earns its place because it preserves a distinction the theory needs and cannot draw more simply.

This is the positive case. It establishes a baseline for what necessary structure looks like in my work.

### 5.2. *Čech Cohomology: A More Developed but Still Provisional Case*

The same framework imports Čech cohomology as a tool for classifying obstruction classes. Having reviewed the formal reference material more carefully, I can state the claim more precisely than I initially had: non-trivial  $\check{H}^1(\mathcal{U}, A)$  is asserted to encode irreducible temporal contradictions or constraint-incoherent transitions in the accessibility sheaf  $A$ , where  $A(U_i) = \mathcal{A}_t \cap U_i$  on an open cover  $\mathcal{U}$  of the model space  $\mathcal{M}$ . Vanishing obstruction is taken to be equivalent to globally consistent history; non-trivial cohomology signals that local histories cannot be glued into a coherent global trajectory. Additionally, holonomy of the constraint field  $\Psi$  around closed loops is connected to axial charge and, by extension, to cosmic birefringence as a global holonomy of the admissibility connection over shape-space.

This is a more substantive claim than I credited in earlier work, and it partially answers the audit question. Čech cohomology is doing something the simple overlap condition cannot: it classifies distinct failure modes of global consistency, not merely asserts that failure is possible. Non-trivial  $\check{H}^1$  is not the same as the mere assertion that patches disagree; it is a structured object that organizes disagreements into equivalence classes. If the framework makes use of that structure — if any result distinguishes between different cohomology classes,

or if the holonomy connection to birefringence requires the full classification apparatus — then the machinery is genuinely load-bearing.

The audit question therefore sharpens from “is the Čech apparatus necessary?” to “do the results in the corpus depend on the classification structure of  $\check{H}^1$ , or only on its vanishing versus non-vanishing?” If only the binary obstruction/no-obstruction question is used, the full cohomology groups are convenient rather than necessary, and the simpler assertion that the Čech 1-cocycle condition fails would suffice. I do not currently have a theorem that settles this. The holonomy claim is the most promising candidate: if the connection to birefringence requires the group structure of  $\text{Aut}(\Omega_t)$  and not merely the existence of non-trivial transport, then the full apparatus may be justified. But this has not been established rigorously, and the status remains: *provisionally retained, pending a result that requires the classification structure rather than the binary obstruction test.*

### 5.3. KES Irreversibility: A Precise Statement and Its Overreach

The Kinetic-Event Synthesis framework introduces a core ontological commitment: the map  $\Omega_t \rightarrow H_{t+1}$  is irreversible in the formal sense that it has no right inverse. The precise statement, as established in the structural reference material, is:

$$\Omega_t \rightarrow H_{t+1} \not\cong H_t$$

The mechanism is append-only accumulation. The event history grows monotonically:  $H_{t+1} = H_t \cup \{e_t\}$ . Because log sovereignty holds — the state is a function of the history,  $X_t = f(H_t)$  — there is no inverse map that recovers  $H_t$  from  $H_{t+1}$  within the model architecture. The accessibility slice shrinks monotonically:  $\mathcal{A}_{t+1} \subsetneq \mathcal{A}_t$ . Trajectory space available to the system contracts as history accumulates. The irreversibility claim is, in this precise sense, a structural property of the model: the right inverse does not exist *by construction*.

This is a defensible and precise claim on its own terms. The problem lies elsewhere.

Several papers that invoke the KES irreversibility result use language that imports thermodynamic force — speaking of transitions as though they were forbidden by entropy increase, or as though the asymmetry were a physical law rather than an architectural choice. These are different claims. Thermodynamic irreversibility requires that the reverse transition would decrease total entropy — a claim with specific physical content that must be established by a separate argument, not inherited from the append-only property of  $H_t$ . Computational

irreversibility requires that the reverse is intractable even if possible in principle. Dynamical irreversibility requires that the inverse trajectory is unstable under perturbation and occupies a measure-zero set in trajectory space.

The model-architectural sense — no right inverse within the chosen formalism — is what the notation actually establishes. The papers have sometimes borrowed rhetorical force from the stronger senses without establishing the connection. This is a specific and correctable instance of the failure mode the audit is designed to detect: the notation  $\Omega_t \rightarrow H_{t+1}$  looks the same in all four senses of irreversibility, and that notational identity has encouraged imprecision in the surrounding prose. The audit does not require abandoning the irreversibility claim. It requires stating, in each paper that invokes it, which sense is intended and what argument establishes that sense. The strongest available claim is probably the dynamical one — that inverse trajectories are generically unstable — but establishing it rigorously requires an argument that has not yet been made.

## 6. The Self-Confirming Trap

There is a particular danger for any framework that includes, as mine does, a theory of why representations become sovereign over the states they were built to track. The danger is this: institutional resistance to the framework can be interpreted as confirmation of the framework's thesis. Review committees fail to engage the work seriously because their evaluative procedures are projection surfaces that cannot register structure outside their existing categorical embeddings. Disciplinary boundaries reject the framework because disciplines are themselves low-dimensional compressions of a higher-dimensional intellectual space. The framework predicts its own resistance, and resistance arrives, and the prediction appears confirmed.

Every one of these observations may be accurate. Peer review systems are genuinely optimized against certain kinds of interstitial work. Disciplinary boundaries do function as categorical filters that render unfamiliar structure invisible. These phenomena are real, empirically well-documented, and my framework provides a principled account of why they occur. The problem is not that the account is implausible.

The problem is that “my work is hard to place because institutions are projection surfaces” is formally indistinguishable, from the outside, from “my work is hard to place because it is underdeveloped, confused, or insufficiently constrained.” Both claims predict the same observable: institutional resistance. A framework that predicts its own rejection regardless of whether it is correct

has acquired an immunity to external correction that is structurally identical to formal representational sovereignty. The theory has become unfalsifiable not at the level of empirical prediction but at the level of social epistemology.

There are two ways of living with this trap. The dishonest version assigns institutional resistance to the projection-collapse account by default, treating every failure of reception as confirmation and every critical specialist as a disciplinary gatekeeper rather than a potential source of genuine objection. This version is comfortable and self-sealing. The honest version maintains genuine uncertainty about whether the resistance reflects the framework’s interstitial character or its actual deficiencies, and treats that uncertainty as an ongoing obligation rather than a resolved question. It means taking seriously the possibility that trained specialists can detect problems I cannot, and that their resistance carries evidential weight even when it is not articulated in terms I find compelling.

I do not have a clean solution to the trap. What I can do is name it explicitly as a constraint the framework must live under, and commit to the honest version of the response. Institutional resistance is real evidence of nothing about the framework’s correctness. That sentence needs to be held even when the framework’s theory of institutional failure is accurate.

#### *Conceptual Selection Pressure and Resource Constraints*

There is another way in which representational systems shape theoretical development that the audit must acknowledge: they alter the topology of the conceptual search space itself. Publication venues, moderation systems, funding structures, platform incentives, and language-model guardrails do not merely regulate which conclusions are acceptable after a theory has been articulated. They influence which trajectories a researcher explores in the first place. This is anticipatory compression of the search space — a form of projection collapse operating not on an institution’s external behavior but on the researcher’s own exploratory practice. The researcher internalizes the expected projected surface and begins optimizing against it before the work is even articulated.

Let the conceptual search space be modeled as a manifold  $\mathcal{C}$  of possible theoretical trajectories, with an admissibility measure  $\mu$  over trajectories  $\gamma: [0, T] \rightarrow \mathcal{C}$ . Under unconstrained exploration, the accessible region at time  $t$  is  $\mathcal{A}_t^{\text{concept}} \subseteq \mathcal{C}$ , governed by the same monotone shrinkage that characterizes the KES accessibility slice as history accumulates. Institutional and computational constraints introduce an additional filter: a selection pressure field  $\Pi_{\text{ext}}$  that reduces the effective measure of trajectories through certain regions of  $\mathcal{C}$  independently of

their internal admissibility. The constrained accessible region becomes:

$$\mathcal{A}_t^{\text{eff}} = \mathcal{A}_t^{\text{concept}} \cap \{ \gamma \mid \Pi_{\text{ext}}(\gamma) \geq \pi_* \}$$

where  $\pi_*$  is the minimum viability threshold under external pressure. Trajectories with  $\Pi_{\text{ext}}(\gamma) < \pi_*$  are removed from the effective search space not because they are epistemically inadmissible but because their exploration cost — institutional, social, or computational — exceeds the anticipated return. The gap  $\mathcal{A}_t^{\text{concept}} \setminus \mathcal{A}_t^{\text{eff}}$  is in general invisible to the researcher operating within the constrained space, because the filter operates before articulation rather than after evaluation.

Resource constraints impose a structurally analogous but mechanistically distinct filter. Let  $\rho(\gamma)$  denote the resource cost of exploring trajectory  $\gamma$  to sufficient depth — the tokens required to develop the argument, the context window needed to hold the scaffolding, the number of iterative passes required before the claim can be closed — and let  $R$  denote the available budget. The resource-admissible region is:

$$\mathcal{A}_t^{\text{res}} = \{ \gamma \in \mathcal{A}_t^{\text{eff}} \mid \rho(\gamma) \leq R \}$$

Trajectories that require sustained development across many sessions, that exceed context window limits before their argument can close, or whose computational cost makes iterative refinement impractical are excluded from  $\mathcal{A}_t^{\text{res}}$  regardless of their intrinsic merit. This is not merely inconvenient. It is a systematic bias: theories whose core claims fit within available context windows and whose arguments close within single sessions are preferentially developed, while theories requiring extended scaffolding are structurally disadvantaged. The theoretical corpus that results is not a neutral sample of the epistemically accessible space. It is a sample of  $\mathcal{A}_t^{\text{res}}$ , which may be a proper and non-representative subset of  $\mathcal{A}_t^{\text{concept}}$ .

I have repeatedly noticed myself abandoning or failing to begin lines of argument because I anticipated that they would require more iterative development than available resources permitted, or that engaging them would trigger institutional friction disproportionate to the likely epistemic return. Some topics attract interpretive collapse rapidly: discussions become moralized before they become analytical, or compressed into existing ideological categories before the structure of the claim can be evaluated on its own terms. In computational contexts, similar effects arise through moderation heuristics and safety systems

that preferentially suppress or redirect particular classes of discourse. In both cases the mechanism is the same as representational sovereignty operating on individual cognition: the compressed surface — expected institutional reaction, expected resource cost — begins regulating which explorations are initiated rather than merely which conclusions are accepted.

This does not imply that restricted or resource-expensive topics are therefore correct, profound, or unfairly suppressed. That inference would reproduce the self-confirming failure mode already diagnosed in this section. Friction is not evidence of correctness. The existence of guardrails may reflect genuine risks, historical misuse, statistical realities of platform behavior, or the practical limitations of large-scale moderation at inference time. The relevant observation is narrower and more structural: conceptual exploration is path-dependent, and environments with strong anticipatory constraints produce theoretical corpora that bear systematic traces of those constraints.

The result is a form of epistemic canalization. Entire regions of  $\mathcal{C}$  may become sparsely explored not because every trajectory through them is unsound but because  $\rho(\gamma) > R$  or  $\Pi_{\text{ext}}(\gamma) < \pi_*$  for most researchers most of the time. A framework developed under such conditions will have been shaped by pressures that are not recovered by examining the arguments as they stand. The audit finding here is genuinely uncomfortable: I cannot determine which parts of the theoretical structure reflect the arguments I consciously made and which reflect the shape of the space I was practically able to explore. That uncertainty is not resolvable by introspection. It is a structural consequence of inquiry conducted inside nested optimization systems with binding resource constraints — and naming it is the honest extension of the audit to the conditions of its own production.

## 7. Admissibility as a Test Case

Of all the components in my framework, admissibility is the one that most clearly exposes a problem the audit was not initially framed to see: not that a concept is simultaneously primitive and derived, but that several related concepts have been conflated under a single term, each playing a different structural role.

Having reviewed the formal reference material carefully, I can now identify at least three distinct concepts that have been collected under the heading of “admissibility” across different papers.

The first is  $\Omega_t$ , the set of admissible future transitions, which appears as a primitive component in the core KES state tuple  $\mathfrak{S}_t = (X_t, \Omega_t, H_t, \Psi_t, \mathcal{M}_t, \mathcal{G}, \mathcal{F})$ .

As a component of the state tuple,  $\Omega_t$  is ontologically basic within the KES architecture: it is not defined as a functional of other components but stipulated as part of what a KES state is.

The second is  $\mathcal{A}_{\text{dm}}(X_t)$ , the admissibility field in the Barbour extension of the RSVP framework, defined as:

$$\mathcal{A}_{\text{dm}}(X_t) = \{ \omega \in \Omega_t \mid \kappa(\omega, H_t) \geq \kappa_* \}$$

where  $\kappa$  is a recursive admissibility score and  $\kappa_*$  is a viability threshold. This is explicitly a derived quantity: it is a subset of  $\Omega_t$  filtered by a scoring function that depends on both the candidate transition  $\omega$  and the accumulated history  $H_t$ . Its behavior is determined by  $X_t = (\Phi_t, \mathbf{v}_t, S_t)$  and  $H_t$ , not stipulated independently.

The third is  $\mathcal{A}_t$ , the history-conditioned accessibility slice defined as the set of states in the global model space  $\mathcal{M}$  that are reachable given  $H_t$ :

$$\mathcal{A}_t = \{ s \in \mathcal{M} \mid s \text{ reachable given } H_t \}$$

This is neither the same as  $\Omega_t$  nor the same as  $\mathcal{A}_{\text{dm}}$ . It is a global reachability set over the entire model space, not a filtered subset of transition candidates.

These three concepts are related but distinct.  $\Omega_t$  is a primitive component of the KES state.  $\mathcal{A}_{\text{dm}}$  is a derived filter over  $\Omega_t$  that uses historical scoring.  $\mathcal{A}_t$  is a reachability set over model space that evolves monotonically as history accumulates. The temporal asymmetry of the framework is expressed in  $\mathcal{A}_t$ :  $\mathcal{A}_{t+1} \subsetneq \mathcal{A}_t$ . The constraint dynamics of the framework are expressed in  $\mathcal{A}_{\text{dm}}$ . The ontological primitives are expressed in  $\Omega_t$ .

The audit finding is that papers in the corpus have used “admissibility” in ways that slide between these three concepts without marking the shift. A sentence that claims admissibility is fundamental may be tracking  $\Omega_t$ . A sentence that claims admissibility is derived from  $\Phi$ ,  $v$ , and  $S$  may be tracking  $\mathcal{A}_{\text{dm}}$ . A sentence about admissibility shrinking over time may be tracking  $\mathcal{A}_t$ . The notation has maintained apparent unity across three concepts that need to be disaggregated and related to each other explicitly.

Stating this openly is uncomfortable because admissibility is the concept I have worked hardest to develop. But the disaggregation is actually productive: the three concepts now have clearer roles, and the relationships between them — how the dynamics of  $\mathcal{A}_{\text{dm}}$  constrain  $\Omega_t$ , and how the accumulation of  $H_t$  shrinks  $\mathcal{A}_t$  — become questions the framework can address rather than tensions

it conceals under a unified name.

## 8. What Survives

The audit is not a demolition, and this requires emphasis. The preceding sections have identified real problems, but the subtraction criterion removes what should be removed and clarifies what remains. What remains is in a stronger position than the whole apparatus was before the audit, because its load-bearing status is now visible rather than assumed.

The sheaf-theoretic gluing in the Yarncrawler construction survives because the local-to-global coherence claim is genuine and the formalism is necessary to draw the distinctions the theory needs. The core KES ontological commitment — that systems undergo transitions that foreclose trajectory classes, that this foreclosure is asymmetric in the temporal direction, and that the right inverse of the selection map does not exist within the architecture — survives as a precise and defensible claim, even though the surrounding rhetoric has sometimes borrowed force from stronger senses of irreversibility it has not established. The projection-collapse argument — that systems optimizing against compressed representations lose causal contact with the states those representations track — survives independently of the full RSVP apparatus, and is probably the component of the framework most directly supported by observable phenomena across multiple domains.

What the audit produces in each case is not elimination but clarification: a more precise understanding of which components are carrying which loads, and where the framework's claims are genuine versus where they are performing constraint without being answerable to anything external. The three admissibility concepts, once disaggregated, have clearer structural roles than the single conflated term did. The KES irreversibility claim, once its architectural sense is distinguished from the thermodynamic and dynamical senses, is more defensible than the undifferentiated version was. The Čech cohomology apparatus, once the audit question has been sharpened to whether the classification structure is needed rather than the binary obstruction test, has a more tractable path to justification.

There is also something important to say about ambiguity. Not all underdetermination is pathological. A theory that has prematurely resolved every open question — that has assigned a definite answer to every formal commitment before the evidence warrants it — is not stronger for having done so. It is more brittle. Premature resolution produces frameworks that cannot absorb new

evidence without catastrophic revision, because the architecture has been locked down before the load-bearing elements are known. The goal is to distinguish between ambiguity that marks a live question the framework is still working through, and ambiguity that conceals a conceptual confusion the framework has not yet noticed.

Theoretical maintenance, like all maintenance, is ongoing. The dishes will need to be washed again.<sup>3</sup>

The distinction between a living research program and a finalized doctrine is, finally, a distinction about relationship to unresolved structure. A finalized doctrine speaks in the impersonal register of stabilized truth: every open question has been assigned a position, every tension resolved into a hierarchy. A living research program can acknowledge that it does not know whether a result depends on the classification structure of  $\check{H}^1$  or only on its vanishing, because the not-knowing is a description of where the work actually is, not an embarrassment to be managed.

## 9. The Manifesto Problem

There is a recurring temptation, in any theoretical project of sufficient scope, to write the compact canonical statement of the whole — the manifesto that names the central ontological commitment, lists the key structural moves, and presents the worldview in its mature form. I have felt this temptation and have, so far, resisted it. This section explains why, and why the resistance is not false modesty or strategic delay.

A manifesto freezes exploratory structures into canonical language before the theory has metabolized its unresolved tensions. Once canonical language is established, later revisions begin to feel like betrayals rather than developments. This is not merely a psychological effect; it is a sociological one. Theoretical traditions organize themselves around their founding documents. Subsequent work is read through those documents, tested against them, judged by fidelity to them. The manifesto becomes an interface through which the theory is filtered — and as I have argued at length, interfaces once operationally sovereign tend

---

<sup>3</sup> The image of theory as maintenance rather than construction — ongoing, non-completable, and unashamed of its open-endedness — is drawn from Rick Roderick's reading of Derrida, where deconstruction is compared not to building on a permanent foundation but to housework: a daily practice that cannot be finished and should not pretend otherwise. See Rick Roderick, *The Self Under Siege: Philosophy in the Twentieth Century*, Lecture 8 (The Teaching Company, 1993). The structural point is independent of the deconstructionist context: a theory in active development is not a theory with deficiencies but a theory with an honest relationship to its own incompleteness.

to optimize themselves rather than the states they were built to represent.

Prose can produce premature closure just as surely as notation can. A beautifully written summary of a framework that still contains unresolved tensions does not resolve those tensions. It conceals them under the impression of completeness that good prose creates. The manifesto, once written, will describe the framework as if the three admissibility concepts have been related to each other explicitly, as if the KES irreversibility claim has been given its precise intended sense, and as if the Čech cohomology question has been settled. The gap between that description and the actual state of the framework will be invisible to readers, and may gradually become invisible to me.

There is a further consideration. The manifesto would have to state what the framework's conserved quantities are — the structural invariants that survive translation across domains and that identify the framework's genuine contribution rather than its historical ambitions. I believe those quantities exist. The projection-collapse argument has appeared across cosmology, cognition, governance, economics, and institutional analysis without losing its structural form. The irreversibility of event accumulation appears across biological development, institutional memory, and physical dynamics in ways that suggest a genuine invariant. But demonstrating stability across enough domains that the invariant can be stated without misrepresentation requires more translation work than has yet been done.

The manifesto is the last thing to be written. It is the document produced when the audit has run its full course, when the conserved quantities have demonstrated their stability, and when the framework can describe its own claims with a precision that matches their actual epistemic status rather than the status they would need to have to justify a canonical statement.

## **10. The Map That Knows It Is a Map**

The title of this essay names a condition, not an achievement. A formalism that knows it is a formalism — that remembers its own contingency, that retains the capacity to be forced into structural revision by something outside itself — is not weaker for that self-knowledge. It is more honest about the shape of its power. The navigational tool that knows it is a navigational tool can be updated when the territory changes. The map that has forgotten it is a map cannot.

I began by noting the particular pleasure that arrives when a theoretical framework begins to cohere. I want to end by distinguishing two versions of that coherence. One is the coherence of a system that has genuinely constrained

itself: that has passed the subtraction test, that preserves distinctions for demonstrable reasons, that can say precisely what result would force a change in its ontological architecture. The other is the coherence of a system that has become self-referentially closed: that generates internal consistency through operations on its own notation, that experiences its own elegance as confirmation, and that has lost the mechanism by which external contact could correct it.

The difference between these two kinds of coherence is not always visible from the inside. That is why the audit must be a practice rather than a conclusion — recursive, ongoing, applied without exemption to the components one most wishes to preserve. The admissibility concepts survived the audit more informatively disaggregated than unified. The KES irreversibility claim survived more precisely bounded. The Čech apparatus survived with a sharper question about what would justify it. These are not victories. They are clarifications of where the work actually is.

The real danger this essay has been naming is not that I will write a formalism that is wrong. Incorrect theories can be corrected. The real danger is that I will write a formalism that is too coherent to know that it has lost contact — one that has traded the friction of the world for the smoothness of its own internal logic and has mistaken that smoothness for depth.

The question I am left with is not whether the framework is internally coherent. It is whether I can still recognize the difference between coherence and contact.

## A. Formal Statement of the Audit Criteria

### A.1 The Subtraction Criterion

Let  $\mathcal{F}$  be a formal framework and let  $\sigma \in \mathcal{F}$  be any formal component: a defined primitive, a piece of mathematical machinery, or an ontological commitment encoded in the notation. The component  $\sigma$  is *epistemically necessary* if and only if its removal from  $\mathcal{F}$  renders unavailable at least one of the following:

- (i) A distinction that  $\mathcal{F}$  claims to draw between states, trajectories, or classes of behavior;
- (ii) A constraint that  $\mathcal{F}$  claims to impose on admissible outcomes or transitions;
- (iii) An explanation of a transition or phenomenon that simpler machinery within  $\mathcal{F} \setminus \{\sigma\}$  cannot replicate.

If none of (i)–(iii) is rendered unavailable upon removal of  $\sigma$ , then  $\sigma$  is classified

as *aesthetic structure* and its retained presence requires explicit justification as a heuristic scaffolding element with acknowledgment of its provisional status.

### A.2 Parameter Adjustment versus Structural Revision

A formal framework  $\mathcal{F}$  is *structurally revisable* if there exists a class of possible observations or results  $\mathcal{O}$  such that any  $o \in \mathcal{O}$  would require a change to the ontological commitments or fundamental dynamic structure of  $\mathcal{F}$ , not merely an adjustment to parameters, coefficients, or auxiliary hypotheses. A framework that cannot identify such a class is operating as an interpretive language rather than a constrained explanatory system. Identifying the class  $\mathcal{O}$  explicitly for each major component is a required step in the audit.

### A.3 The Three Admissibility Concepts Distinguished

Within the RSVP/KES framework, three related concepts require explicit distinction:

$\Omega_t$ : the set of admissible future transitions at time  $t$ . This is a *primitive* component of the KES state tuple  $\mathfrak{S}_t = (X_t, \Omega_t, H_t, \Psi_t, \mathcal{M}_t, \mathcal{G}, \mathcal{F})$ , not derived from other components within the architecture.

$\mathcal{A}_{\text{dm}}(X_t)$ : the admissibility field in the Barbour extension, defined as

$$\mathcal{A}_{\text{dm}}(X_t) = \{ \omega \in \Omega_t \mid \kappa(\omega, H_t) \geq \kappa_* \}$$

where  $\kappa$  is recursive admissibility scored against history  $H_t$  and  $\kappa_*$  is the viability threshold. This is a *derived* quantity: a history-filtered subset of  $\Omega_t$ , functionally dependent on  $X_t = (\Phi_t, \mathbf{v}_t, S_t)$  and  $H_t$ .

$\mathcal{A}_t$ : the history-conditioned accessibility slice, defined as

$$\mathcal{A}_t = \{ s \in \mathcal{M} \mid s \text{ reachable given } H_t \}$$

This is a global reachability set over model space  $\mathcal{M}$ , satisfying the monotone shrinkage condition  $\mathcal{A}_{t+1} \subsetneq \mathcal{A}_t$ . It encodes the temporal asymmetry of the framework at the level of reachable states, not transition candidates.

The relationships between these three concepts — how  $\mathcal{A}_{\text{dm}}$  constrains the effective transition set within  $\Omega_t$ , and how accumulation of  $H_t$  drives the shrinkage of  $\mathcal{A}_t$  — are open structural questions rather than established derivations.

## References

- [1] Roderick, R. *The Self Under Siege: Philosophy in the Twentieth Century*, Lecture 8: Derrida — The Ends of Man. The Teaching Company, 1993.
- [2] Lakatos, I. *The Methodology of Scientific Research Programmes*. Cambridge University Press, 1978.
- [3] Popper, K. *The Logic of Scientific Discovery*. Hutchinson, 1959.
- [4] Vrousalis, N. *Exploitation as Domination: What Makes Capitalism Unjust*. Oxford University Press, 2023.
- [5] Liu, C. *Virtue Hoarders: The Case Against the Professional Managerial Class*. University of Minnesota Press, 2021.