Beyond Prediction Error: Toward an Ecological and Topological Theory of Cognition

Flyxion

Independent Researcher

May 2026

Contents

Introduction

The dominant frameworks of contemporary cognitive neuroscience and artificial intelligence share a common presupposition that is almost never made explicit: that cognition is, at its core, a process of prediction-error minimization. Whether the formalism is predictive coding, active inference, Bayesian brain theory, or the free-energy principle, the underlying architecture is consistent. The organism receives sensory signals, compares them against internally generated forecasts, and updates its model to reduce discrepancy. Intelligence, on this account, is the capacity to be accurately surprised — to maintain an internal generative model whose predictions correspond closely enough to the world that the organism can navigate it without catastrophic failure.

This essay argues that this presupposition, while technically useful in restricted domains, has been systematically inflated beyond its legitimate scope. The conflation of analytical description with biological ontology has produced a neuroscience that frequently mistakes the observer's mathematical tools for the causal architecture of the organism itself. The organism has been remodeled in the image of the modeler's formalism, and what began as a productive metaphor has hardened into a reification.

The central thesis of this essay is that cognition is not fundamentally prediction-error minimization but the maintenance of admissible trajectories within dynamic ecological fields. Two distinctions are fundamental to everything that follows:

$$\text{Prediction} \neq \text{Agency}$$

and

$$\text{Compression} \neq \text{Organization}.$$

These inequalities are not obvious within the dominant paradigm. Indeed, much of the rhetorical force of predictive-processing frameworks derives from their apparent ability to absorb the concepts of agency and organization into the mathematical machinery of prediction. This essay attempts to show why that absorption is illusory, why it erases features of biological organization that are explanatorily essential, and what an alternative framework — grounded in ecological psychology, process philosophy, embodied cognition, and admissibility-theoretic accounts of trajectory navigation — must preserve in order to succeed.

The argument proceeds through several interconnected movements. The first movement traces the historical path by which computational and informational metaphors became reified into claims about neural ontology, culminating in the contemporary dominance of predictive-processing accounts. The second examines Romain Brette's detailed critique of predictive coding's equivocation between signal reconstruction and genuine anticipation, arguing that this equivocation is not merely terminological but reveals a deep structural confusion about what biological systems are actually doing. The third develops an alternative framework centered on admissibility, trajectory continuity, and organizational persistence. The fourth addresses a cluster of clarifications and extensions that the admissibility framework requires: the proper subordinate role of local prediction, the distinction between correspondence and coherence, a geometric account of cognitive pathology, the multi-scale temporal layering of persistence, and the mechanism by which prediction systems simulate agency without instantiating it. The fifth examines symbolic structures — essays, notes, interfaces, institutions — not as memory stores but as topological stabilization operators that preserve coherence across otherwise volatile transformations. The sixth develops the full topological ontology of cognition: stability basins, organizational folds, and boundary management, then extends the framework through the formal analogy between Lie group local-global structure and the relationship between local predictive operators and global admissibility topology, including the $\mathfrak{sl}_2$ seed structure and its polynomial realization, before formalizing the stratified decomposition of admissibility that captures singular organizational boundaries and pathological stratum descent, and sharpening the distinction between optimization as loss minimization and admissibility as categorical constraint maintenance. The chapter closes with a differentiation of the admissibility framework from classical enactivism and a speculative account of empirical formalisms. The seventh draws implications for the theory of agency and the structural limits of current artificial intelligence.

Throughout, the essay resists the temptation to replace one dogma with another. Prediction obviously occurs. Organisms exploit regularities. Compression is cognitively real. The question is whether these phenomena constitute the primary ontological principle of cognition, or whether they are better understood as subordinate processes within a more fundamental organization whose central problem is not epistemic accuracy but viable persistence.

Chapter 1

The Computational Metaphor and the Reification of Models

1.1 From Cybernetics to Predictive Processing

The history of cognitive science is substantially a history of borrowed metaphors. Every major technical innovation in information processing has produced a corresponding wave of claims about the nature of the mind. The telegraph suggested that the nervous system was a communications network. The programmable digital computer suggested that the brain was a symbol-manipulation device executing stored programs. Bayesian statistics suggested that inference was the brain's fundamental operation. Each of these borrowings was genuinely productive at the level of generating testable models and explanatory frameworks. Each was also, in ways that took decades to clarify, a category error — a confusion between the structure of the analytical tool and the structure of the phenomenon being analyzed.

Cybernetics, the earliest systematic attempt to unify cognitive and mechanical accounts of purposive behavior, introduced the negative feedback loop as a candidate ontological primitive for goal-directed action [11]. The organism, on this account, is a system that measures the discrepancy between its current state and a target state, and generates corrective outputs proportional to that discrepancy. This formalism was enormously powerful and genuinely captured something important about homeostatic regulation. Its extrapolation to cognition in general, however, required the assumption that cognitive goals were specifiable in terms of representable target states — an assumption that was built in from the beginning and rarely examined.

The computationalist program that followed extended this architecture into the domain of symbolic manipulation. Minds were characterized as Turing machines operating over internal representations, with perception as input transduction, memory as storage, and reasoning as rule-governed symbol manipulation [9, 10]. The theoretical vocabulary of this tradition — representations, rules, programs, execution, registers — was imported wholesale from computer science, and with it came an implicit commitment to a particular ontology: that cognition is constituted by the manipulation of discrete, context-independent symbolic structures.

Connectionism offered a partial correction by emphasizing distributed, subsymbolic, and statistically grounded processing, but it preserved the core representationalist commitment. The brain was still a device for computing mappings from input to output, where the mappings were learned rather than programmed. The shift from serial symbol manipulation to parallel distributed processing changed the implementation story without changing the fundamental explanatory structure.

The contemporary dominance of predictive processing represents the latest and most sophisticated iteration of this lineage. Drawing on Helmholtz's notion of perception as unconscious inference and integrating Bayesian probability theory, hierarchical generative models, and variational inference, frameworks such as predictive coding [8, 6] and active inference [7] propose that the brain is fundamentally a prediction machine. On the canonical account, the brain maintains a hierarchical generative model of the causes of sensory data. At each level of the hierarchy, predictions flow downward and prediction errors flow upward. The system learns by adjusting its generative model to minimize prediction errors across the hierarchy, and it acts by generating predictions about the sensory consequences of action and then fulfilling those predictions through motor behavior. In the active inference framework, this architecture subsumes both perception and action under a unified imperative to minimize free energy — a measure of the divergence between the organism's generative model and the actual distribution of sensory data.

The theoretical elegance of this account is genuine and its empirical productivity is substantial. It has generated novel predictions about perceptual phenomena, neural anatomy, and clinical conditions, and it has provided a common framework for work that was previously theoretically fragmented. Nevertheless, the present essay argues that this elegance comes at a cost: the cost of systematically conflating what organisms do with how scientists model what organisms do.

1.2 The Observer's Tools Become Ontology

The deep problem with predictive-processing frameworks is not that they are mathematically incorrect or empirically false. It is that they commit what might be called the reification error: they treat the mathematical tools of the observer as constitutive features of the phenomenon being observed. Neuroscience has increasingly inherited this error from cognitive science, and it has produced a discipline that is, in a precise sense, a high-level autobiography of its own methodology.

Consider the structure of the error formally. Let X denote the full space of ecological trajectories available to an organism — the complete manifold of dynamical paths through its environment, including all the temporal, metabolic, embodied, and relational dimensions of its existence. A scientist studying this organism does not have access to X directly. The scientist constructs a compressed representational manifold

$$M = \pi(X),$$

where $\pi : X \to M$ is a projection that retains whatever features the scientist's observational apparatus and theoretical commitments make salient. The scientist then identifies correlations between M and external variables, constructs latent-variable models, applies Bayesian inference over compressed descriptions, and generates predictions about the organism's behavior in terms of structures within M.

None of this is illegitimate as a scientific practice. The error arises when the scientist concludes that because M is an effective predictive tool, the organism must literally implement the structures of M as its own internal ontology. This inference is invalid. The effectiveness of M as a scientific description does not entail that the organism operates by constructing M internally and reasoning over it. A Fourier transform provides an extraordinarily effective decomposition of a sound wave without thereby becoming the causal architecture of the vibrating string that produced it. Similarly, Bayesian inference over a latent-variable model may predict neural firing patterns without implying that the nervous system implements Bayesian inference.

Predictive-processing frameworks are particularly vulnerable to this error because they are built from mathematical objects — generative models, prediction errors, prior distributions, variational free energy — that naturally invite interpretation as internal computational structures. The formalism invites the neuroscientist to locate these objects inside the organism and treat the organism's behavior as the output of computations over them. But this interpretive move is precisely where the map begins to eat the territory.

The concept of a "neural code" exemplifies the problem. When neuroscientists observe correlations between neural firing patterns and external variables, they describe the neurons as "encoding" those variables. But correlation between a physical quantity and an external variable does not imply encoding in any computationally significant sense. The correlation is a feature of the scientist's description, not necessarily a feature of the organism's causal organization. As Brette argues at length, the encoding framing imports from outside the organism precisely the interpretive machinery that is supposed to be explained [1]. The scientist already knows what the "stimulus" is; the organism does not have this privileged external perspective. Treating neural activity as encoding a stimulus therefore presupposes, rather than explains, how the organism relates to its environment.

Definition 1.1. *The* reification error *is the inference from the effectiveness of a mathematical description $M = \pi(X)$ to the conclusion that the system being described literally implements the structures of M as its causal architecture.*

Proposition 1.2. *The reification error is not merely philosophical but has empirical consequences: it generates research programs whose questions are systematically ill-posed, because they ask what internal representational structures the organism uses to implement operations that the scientist has already performed externally on behalf of the organism.*

Proof. Let $\mathcal{Q}$ be the class of research questions generated by the assumption that the organism implements M internally. A question $q \in \mathcal{Q}$ takes the form: "What is the neural mechanism that computes the transformation $f : M \to M'$?" But if f was defined by the scientist as an operation on $M = \pi(X)$, and if the organism does not have access to M as a self-standing representational object, then the organism cannot literally compute f. The question therefore presupposes a computational architecture that may not exist. Research programs organized around $\mathcal{Q}$ will produce results that are accurate at the level of M — i.e., they will find neural correlates of whatever the scientist chose to measure — but will systematically misattribute causal structure because they are searching for implementations of operations defined externally. This is not a contingent methodological failure but a structural consequence of the reification error. □

1.3 Epistemic Phlogiston

The concept of "information processing" is perhaps the clearest instance of what this essay will call epistemic phlogiston: an explanatory placeholder that inherits its apparent meaning from the interpreter's external knowledge rather than from any intrinsic property of the system being described. Phlogiston was a substance postulated to explain combustion; it had all the theoretical hallmarks of a genuine explanation — it unified a range of phenomena, generated predictions, and was embedded in a coherent theoretical network — but it turned out to explain nothing because it was defined circularly in terms of the very phenomenon it was supposed to account for.

Information processing plays a similar role in cognitive neuroscience. When a neuroscientist says that a neural area "processes spatial information" or that a circuit "represents temporal regularities," the content of these claims derives not from properties of the neural activity itself but from the scientist's prior knowledge of what spatial and temporal structures are relevant to the organism's task. The neural activity is real; the "information" it processes is a projection of the scientist's interpretive framework onto that activity. The claim appears explanatory because it uses the vocabulary of computation, but it actually shifts the explanatory burden into a black box labeled "information" whose contents are supplied from outside.

This is not a claim that neural systems do nothing. It is a claim that describing them in terms of information, representation, and coding frequently delays rather than advances genuine explanation. What would a genuine explanation look like? It would specify the dynamical, metabolic, and ecological processes by which the neural system participates in maintaining the organism's viability — the actual physical mechanisms by which neural activity is coupled to environmental structure and organismic organization, rather than the observer's compressed description of that coupling.

Brette extends this critique to the concept of "neural coding" specifically, arguing that the language of codes and signals imports an entire communications-theoretic framework that was developed for engineering systems in which a sender, a channel, and a receiver are independently specifiable [2]. In the nervous system, the “signal” and the “receiver” are not independent; the same neural circuits that “receive” sensory input also generate the motor commands that transform the sensory environment, participate in metabolic regulation, and are continuously modulated by internal state. Applying communications-theoretic vocabulary to this system systematically misrepresents its organizational structure by treating as separate what is fundamentally coupled.

The same critique applies to the concept of “representations.” A representation, in the logico-computational sense, is a structure that stands in for something else by virtue of a systematic mapping relation. For this concept to be explanatorily useful, it must be possible to specify the mapping independently of the system that is supposed to use the representation. But in biological systems, there is no such independent specification. The neural structures that are supposed to implement representations are themselves causally embedded in the processes that constitute the organism’s interaction with its environment. To say that these structures “represent” features of the environment is, again, to project the scientist’s interpretive framework onto a coupled dynamical system in a way that obscures rather than illuminates the causal organization of that system.

Chapter 2

Predictive Coding and the Collapse of Categories

2.1 What Predictive Coding Actually Predicts

Predictive coding, in its original and technically precise formulation, is a signal processing technique in which a sender transmits not a signal itself but the difference between the signal and a locally generated prediction of the signal. This approach reduces transmission costs when the signal is sufficiently regular that local predictions are accurate. The technique was developed for telecommunications and image compression and subsequently applied to the visual system by Rao and Ballard [8], who proposed that the hierarchical organization of visual cortex could be understood as implementing a form of predictive coding in which higher cortical areas generate predictions of lower-area activity and the lower areas transmit only the residual prediction error.

In this technical sense, predictive coding is a specific and testable hypothesis about neural architecture. The hypothesis is interesting and has generated a productive empirical research program. The problem is that as the framework was generalized from visual cortex to the entire brain, and from neural architecture to a theory of cognition, the meaning of "prediction" shifted in ways that were rarely made explicit.

The shift can be described precisely. In the original sense, a prediction is a numerical estimate of a future or concurrent signal value:

$$\hat{x}_{t+1} \approx x_{t+1},$$

where $\hat{x}_{t+1}$ is the predicted signal and x_{t+1} is the actual signal. A prediction error is the scalar or vector difference $\epsilon_{t+1} = x_{t+1} - \hat{x}_{t+1}$. These are well-defined mathematical objects, and their neural correlates can in principle be empirically identified.

As the framework expanded, however, "prediction" came to encompass a much broader range of phenomena: perceptual anticipation, goal-directed action, planning, decision-making, and ultimately all adaptive behavior. Organisms are said to "predict" their sensory inputs, to "predict" the outcomes of their actions, to maintain "prior predictions" about the statistical structure of their environments, and to minimize "prediction error" as the fundamental imperative of all neural processing. In this expanded sense, the technical precision of the original formulation has been dissolved into a diffuse theoretical vocabulary that inherits its intuitive force from ordinary-language notions of anticipation, expectation, and foresight.

Brette's central contribution is to demonstrate that this expansion involves not a legitimate generalization but an equivocation [2]. The expanded framework uses the vocabulary of the technical sense while smuggling in the intuitive content of the ordinary-language sense, thereby creating the appearance that the technical framework explains phenomena that it merely redescribes. The result is a framework that appears far more powerful than it is, because its apparent scope derives from conceptual ambiguity rather than theoretical depth.

2.2 Prediction Versus Anticipation

The distinction between prediction and anticipation is philosophically fundamental and has been systematically obscured by the expansion of predictive processing frameworks. To make the distinction precise, it is necessary to examine concrete cases.

Consider a runner in starting blocks. As the starter raises the pistol, the runner's body enters a state of heightened readiness: muscles partially contracted, weight shifted forward, attention focused. This state constitutes genuine anticipation of the starting signal. But it is not a prediction in the technical sense. The runner is not generating a numerical estimate of when the signal will occur or what sensory stimulus it will produce. Rather, the runner's body is configured to exploit a regularized environmental event — the starting signal — in a way that is conditional, action-relative, and ecologically embedded. The anticipation is structured by the runner's practical relationship to the race, not by any internal probability distribution over sensory inputs.

Consider, further, someone who sees clouds gathering and retrieves an umbrella before leaving the house. This too is genuine anticipation: the person is organizing their present behavior in relation to a probable future state of the environment. Again, however, the anticipation is not a numerical prediction of future sensory inputs. It is a practical reorganization of current behavior in light of an ecologically relevant regularity — the correlation between cloud formations and rainfall — that has been internalized not as an explicit probabilistic model but as a disposition to act in certain ways in certain circumstances.

What these cases share, and what distinguishes them from technical prediction, is that they are fundamentally action-relative and constraint-sensitive. Anticipation, in the ecologically significant sense, is not the generation of predictions about future sensory states but the configuration of present organizational states to exploit environmental regularities in ways that preserve or extend the organism's admissible action-space. Brette captures this with the formulation that "anticipation means exploiting regularities, not necessarily uttering prophecies" [2].

The distinction can be formalized as follows. Technical prediction is the operation:

$$\text{Prediction} : \text{State}_t \to \widehat{\text{Signal}}_{t+1}.$$

This is a mapping from current state to a numerical estimate of a future sensory datum. Ecological anticipation, by contrast, is the operation:

$$\text{Anticipation} : \text{State}_t \times \text{Regularities} \to \text{Configuration}_{t+\delta},$$

where $\text{Configuration}_{t+\delta}$ denotes an organizational state structured to maintain or expand admissible trajectories given the probabilistic structure of the environment. Anticipation is not primarily an epistemic operation but an organizational one. It does not estimate future inputs; it reorganizes present configurations to remain viable within a structured environment.

This distinction matters enormously for the theory of cognition. If cognition is fundamentally anticipatory in the ecological sense, then what is being managed is not epistemic uncertainty about future sensory data but organizational coherence within a dynamically structured environment. The organism is not primarily trying to be right; it is primarily trying to remain viable.

Definition 2.1. Ecological anticipation *is the dynamical reconfiguration of an organism's organizational state in ways that exploit environmental regularities to maintain or extend the organism's admissible trajectory space. It is to be distinguished from* technical prediction, *which is the generation of numerical estimates of future signal values.*

The significance of this distinction extends to the theory of perception. Predictive processing accounts of perception describe it as a process of inference: the brain uses its generative model to generate predictions about the causes of sensory data, and perception is the result of this inference process. But ecological accounts of perception, following Gibson [3], describe it as a process of affordance detection: the organism perceives its environment directly in terms of the action-possibilities it affords, without the intermediary of inference over representational models.

On the Gibsonian account, when a person sees a chair, they do not first compute an abstract "chair-representation" and then derive from it the action-possibility of sitting. The chair is perceived directly as sit-able; the affordance is given in perception, not derived from it. This is not merely a claim about the speed of processing but a claim about the structure of perceptual experience: its primary content is action-possibility, not sensory qualities organized into object-representations.

This account is deeply incompatible with predictive processing frameworks in which perception is the endpoint of an inferential process over a generative model of the world. Affordance perception is not an inference to the best explanation of sensory data; it is a direct responsiveness to the ecological structure of the environment mediated by the organism's specific action-capacities. The organism is not a hypothesis-testing machine; it is an agent embedded in an action-field.

2.3 The Collapse of Constraint, Preference, and Goal

The active inference extension of predictive processing raises additional problems that go beyond the prediction-versus-anticipation distinction. Active inference proposes that action is not merely the outcome of inference about the world but is itself a form of prediction: the organism acts by generating predictions about the sensory consequences of its intended actions and then fulfilling those predictions through motor commands. This subsumes action under the same mathematical framework as perception, unifying them under the imperative to minimize variational free energy.

This unification has a significant theoretical cost: it requires collapsing the distinction between what the organism predicts and what the organism wants. In active inference frameworks, goals and preferences are encoded as prior probability distributions over sensory states. The organism acts to bring about states that match its priors, which are interpreted as "predictions" about the organism's typical or preferred sensory experience. But this interpretation creates a deep ambiguity: are priors descriptions of what the organism expects, or are they prescriptions of what the organism wants? In active inference, they must be both, which means that prediction and desire are definitionally identical within the framework.

The consequences of this identification are severe. If a runner "predicts" both winning the race and experiencing rain on the way home, the framework has no principled basis for distinguishing between the predictions that drive action and the predictions that merely reflect environmental expectations. Additional theoretical machinery — hierarchical priors, temporal depth, precision-weighting — must be introduced to recover the distinction between goal-directed action and passive environmental modeling, but this machinery is not internal to the basic active inference formalism. It is added ad hoc to recover distinctions that a more adequate ontology would have preserved from the beginning.

More fundamentally, the active inference framework conflates several categories that must be kept distinct for an adequate theory of agency:

A *constraint* is a physical or organizational limit on the organism's possible trajectories — the metabolic requirements for continued existence, the biomechanical limits of movement, the ecological conditions for viable activity.

A *preference* is a dispositional orientation toward certain states or trajectories within the admissible region — what the organism tends toward when unconstrained.

A *goal* is a specific state or trajectory that an organism represents as a target of directed activity.

A *prediction* is an estimate of what state or trajectory will obtain given the current dynamical evolution of the system.

An *admissible trajectory* is a path through the organism's state space that satisfies the organism's constraints and preserves its organizational integrity.

These categories are importantly different, and collapsing them into the unified currency of “prediction” or “free-energy minimization” erases distinctions that are essential for understanding the structure of purposive behavior. A living system cannot be understood solely through scalar loss minimization over a generative model, because the organization of living systems involves the maintenance of constraint relationships, preference structures, and goal-directedness in ways that are not reducible to the minimization of any single scalar quantity.

Theorem 2.2. *No scalar loss function over sensory predictions can fully characterize the organizational structure of a living system.*

Proof. Let $L : \mathcal{S} \to \mathbb{R}$ be a scalar loss function over sensory states $s \in \mathcal{S}$. For L to characterize the organism's organizational imperatives, it must simultaneously encode: (i) hard constraints $\mathcal{C} \subset \mathcal{S}$ on viable states (outside which the organism ceases to exist as an organized system); (ii) preference orderings over states within $\mathcal{C}$; and (iii) goal representations specifying specific target states. But hard constraints are not expressible as finite-valued functions: they require infinite penalty outside $\mathcal{C}$, which makes L discontinuous and prevents gradient-based minimization from respecting them in principle. Furthermore, preference orderings and goal representations are distinct intentional structures that can conflict — the organism may prefer a state it cannot reach given its constraints, or represent as a goal a state it does not prefer under its generic preferences. A single scalar L cannot simultaneously encode all three structures without either collapsing their distinctness or becoming underdetermined. Therefore, no scalar loss function over sensory predictions fully characterizes the organizational structure of a living system. □

Chapter 3

Admissibility and Ecological Organization

3.1 Trajectory Spaces and Viability

An alternative to prediction-error frameworks can be developed from the concept of admissibility. Rather than asking what the organism predicts, we ask what trajectories remain open to it — what paths through its state space preserve its organizational integrity and maintain its continued existence as a viable system.

Definition 3.1. *Let X be the ecological trajectory space of an organism: the manifold of all dynamically possible paths through the organism's state space, including its internal states, behavioral options, and environmental configurations. The* admissible subspace *at time t is the set*

$$\mathcal{A}(t) \subset X$$

of all trajectories in X that are consistent with the organism's physical constraints, metabolic viability, and organizational integrity. A trajectory $\gamma \in X$ is admissible if and only if $\gamma \in \mathcal{A}(t)$ for all t in the relevant temporal interval.

On this framework, cognition is not fundamentally the minimization of prediction error but the navigation of $\mathcal{A}(t)$: the ongoing management of the organism's position within its admissible subspace in ways that preserve viability and extend the range of future admissible trajectories. The organism's central problem is not epistemic — it is not primarily trying to construct an accurate model of the world — but organizational: it is trying to remain inside $\mathcal{A}(t)$ as t advances and as $\mathcal{A}(t)$

itself evolves in response to environmental change.

A question of normativity arises here that the formalism alone does not settle: if a trajectory satisfying $\gamma(t) \in \mathcal{A}(t)$ is by definition admissible, does the framework imply that remaining admissible is always success, and that wider admissible spaces are always better? The answer is that admissibility is a necessary but not sufficient condition for what might be called flourishing. The framework describes the organizational floor — the conditions under which purposive activity remains possible at all — without prescribing which trajectories within the admissible region are most valuable. The normative question of which admissible trajectories to pursue is a separate question from the organizational question of how to remain admissible. The present essay is primarily concerned with the latter, which is prior to the former: one cannot pursue valuable trajectories from outside the admissible region. The claim that extending the range of future admissible trajectories is organizationally desirable is therefore a structural rather than a terminal normative claim — it says that preserving optionality is a condition of continued agency, not that all options are equally worth preserving.

This reframing has several important consequences. First, it relocates the primary explanatory burden from internal representational structures to the organism-environment interface. What matters is not the accuracy of the organism's internal model but the maintenance of its organizational coupling to environmental regularities. Second, it makes the temporal structure of cognition central rather than peripheral. $\mathcal{A}(t)$ is not static; it evolves as the organism acts and as its environment changes. Cognition is the ongoing management of this evolution, not the construction of a timeless world-model. Third, it preserves the reality of prediction and inference as subordinate processes within a larger organizational architecture. Organisms do generate predictions and perform inference, but these operations serve the more fundamental imperative of organizational persistence rather than constituting that imperative.

Proposition 3.2. *The admissibility framework subsumes prediction-error minimization as a special case while preserving the categorical distinctions that prediction-error frameworks collapse.*

Proof. Let $P : X \to \mathbb{R}^n$ be a prediction function mapping trajectories to predicted signal values, and let $E(t) = \|P(\gamma(t)) - x(t)\|$ be the prediction error at time t. Prediction-error minimization corresponds to gradient descent on $E(t)$ over the space of possible trajectories. Within the admissibility framework, this operation is admissible if and only if the gradient descent trajectory remains within $\mathcal{A}(t)$: i.e., if prediction-error minimization does not violate the organism's organizational constraints. When $\mathcal{A}(t) = X$ (no constraints), prediction-error minimization is fully admissible and the frameworks coincide. When $\mathcal{A}(t) \subsetneq X$ (the generic biological case), prediction-error minimization may exit $\mathcal{A}(t)$, in which case the organism must prioritize constraint maintenance over error minimization. The admissibility framework thus subsumes the prediction-error framework while preserving the distinctions between constraints, preferences, and goals that the prediction-error framework collapses into the scalar $E(t)$. □

3.2 Living Systems as Persistence Structures

The central insight of the admissibility framework is that organisms are fundamentally persistence structures rather than inference machines. Their primary organizational imperative is not to reduce uncertainty about the world but to maintain the dynamical conditions under which they can continue to exist as organized systems. This imperative is prior to and more fundamental than any epistemic operation, including prediction, inference, and model construction.

This claim can be sharpened by considering the contrast between predictive accuracy and organizational persistence. A thermostat predicts, in a weak sense, the temperature of the room and generates corrections to maintain it within a target range. An LLM predicts, in a sophisticated sense, the continuation of a text sequence. Neither thermostat nor LLM is a persistence structure in the biological sense. Neither must maintain metabolic continuity, preserve the conditions of its own self-repair, maintain ecological coupling with an environment whose structure it depends on for its continued functioning, or manage the multi-scale organizational hierarchy that constitutes a living body. Their "predictions" are computationally impressive but ontologically shallow: they optimize correspondence rather than persistence.

Living organisms, by contrast, exhibit what we might call *homeodynamic* organization: they maintain their organizational structure not by achieving a static equilibrium but by continuously cycling through controlled perturbation and recovery, actively creating and maintaining the conditions of their own continued existence. This is not a trivially different kind of prediction-error minimization; it is a categorically different organizational principle.

Definition 3.3. *A system S is a* persistence structure *if it maintains the conditions* *of its own continued organization through dynamical coupling with its environment, metabolic self-repair, and active management of its own boundary conditions, such that the primary criterion of success is continued organizational coherence rather than predictive accuracy.*

The distinction between optimization of correspondence and optimization of persistence produces different explanatory frameworks for understanding cognitive pathology. Predictive-processing accounts of conditions such as depression, psychosis, or anxiety characterize them as disorders of prediction or inference: the generative model has inappropriate priors, or the precision-weighting of prediction errors is dysregulated. These accounts are not without insight. But they systematically miss the organizational dimension of these conditions: the collapse of the organism's capacity to maintain viable trajectories through its social and ecological environment, the disintegration of the temporal coherence that makes purposive action possible, the narrowing of the admissible trajectory space to a point where the organism can no longer sustain the conditions of its own vitality.

An admissibility framework makes these organizational dimensions central rather than peripheral. Cognitive pathology is not primarily a disorder of inference but a disorder of trajectory maintenance: a condition in which the organism's capacity to navigate its admissible subspace has been compromised, whether by environmental constraint, metabolic disruption, or the collapse of the stabilizing structures that normally anchor its organizational coherence.

3.3 Entrained Anticipation

The cyanobacterium provides one of the most philosophically important examples in the entire debate about prediction and anticipation. Before dawn, cyanobacteria begin decompacting their chromosomes in preparation for the photosynthetic activity that will begin when sunlight arrives. This is genuine anticipation: the organism is reorganizing its internal state in advance of an environmental event, in a way that is adaptive and functional. Yet it is not prediction in any computationally significant sense. The chromosome does not contain an internal model of sunrise. The cyanobacterium does not infer from sensory evidence that sunrise is imminent and update a probability distribution over future states. The anticipatory organization is embedded in the dynamical coupling between the organism's biochemical rhythms and the environmental periodicity of light and dark.

This is what Brette calls anticipation without representation [2]: the organism exploits regularities not by constructing internal models of them but by being dynamically entrained to them. The anticipatory structure is not a computational operation performed on representations but a physical fact about the organism's embedding in a periodic environment. The organism's internal dynamics are shaped by the environment's periodicities in a way that produces adaptive reorganization without requiring any explicit predictive machinery.

The concept of entrainment is crucial here. An entrained system is one whose internal dynamics have become coupled to the periodicities of an external driving system in a way that causes the internal dynamics to anticipate the driving system's future states. This anticipation is not inference-based but dynamical: it is a consequence of the physical coupling between two oscillatory systems, not the output of an inference algorithm operating over sensory data.

Definition 3.4. Anticipatory entrainment *is the condition in which an organism's internal dynamical organization is physically coupled to environmental periodicities in a way that produces adaptive reorganization in advance of environmental events, without requiring the construction or manipulation of explicit predictive representations.*

Anticipatory entrainment is not a marginal phenomenon. It is pervasive in biology at all scales, from the circadian rhythms of single cells to the cultural and institutional rhythms of human societies. Human cognition is shot through with entrained anticipation: we wake before our alarm, complete familiar sentences before they are uttered, flinch before the loud noise reaches full amplitude. These are not instances of sophisticated Bayesian inference; they are instances of dynamical coupling between an organism and its structured environment.

The contrast between anticipatory entrainment and technical prediction reveals a fundamental difference in the kind of system that exhibits each. Technical prediction requires a system with an internal model: a representation of the environment's statistical structure, together with computational machinery for using that representation to generate estimates of future states. Anticipatory entrainment requires a system that is physically coupled to the environment's structure in a way that its internal dynamics partially mirror the environment's dynamics. The first is an epistemic relationship between model and world; the second is a physical relationship between coupled dynamical systems.

This distinction has deep consequences for the theory of artificial intelligence. Current AI systems are, almost exclusively, prediction systems in the technical sense: they construct internal representations of statistical regularities in training data and use those representations to generate continuations, completions, or classifications. They are not, in any meaningful sense, entrained to their environment, because they have no physical body, no metabolic organization, and no ecological embedding that could support genuine entrainment. Their apparent anticipatory competence is a consequence of sophisticated statistical modeling, not dynamical coupling.

Chapter 4

Clarifications and Extensions

4.1 Prediction as Local Regulation Rather Than Global Ontology

One possible misunderstanding of the present framework is that it denies the existence or importance of predictive processes altogether. This is not the claim. Organisms clearly exploit temporal regularities, generate expectations, and perform local anticipatory adjustments. The argument developed here is instead that predictive operations are subordinate regulatory mechanisms within a larger organizational structure whose primary imperative is persistence rather than correspondence.

A useful analogy is biological homeostasis. A thermostat-like regulation loop may maintain body temperature within a narrow range, but the organism itself is not reducible to temperature regulation. Temperature regulation is one local stabilizing operation embedded within a multi-scale organizational structure whose overall imperative is viability. Similarly, predictive mechanisms may participate in local stabilization without constituting the ontological foundation of cognition itself.

Formally, let

$$P_i : X \to \mathbb{R}^{n_i}$$

denote a family of local predictive operators acting over restricted subspaces of the organism's trajectory manifold. These operators generate local anticipatory adjustments:

$$P_i(x_t) \mapsto \hat{x}_{t+\delta}.$$

The admissibility framework does not reject these operations. It embeds them within a higher-order organizational constraint:

$$\gamma(t) \in \mathcal{A}(t),$$

where the preservation of admissibility takes precedence over minimizing any individual predictive discrepancy.

This distinction matters because predictive optimization can become locally successful while globally destructive. An organism may reduce uncertainty within a restricted domain while simultaneously collapsing its larger admissible trajectory space. Addiction, compulsive behavior, bureaucratic rigidity, and pathological optimization systems all exhibit this pattern: local stabilization produces global organizational failure. Prediction therefore remains real but non-sovereign. It is one stabilization strategy among many within a broader ecology of persistence.

4.2 The Distinction Between Correspondence and Coherence

A second ambiguity concerns the concepts of correspondence and coherence, which are frequently treated as interchangeable within computational and representational frameworks but refer to fundamentally different organizational relationships.

Correspondence concerns representational alignment: a representation $R(x)$ accurately tracks some feature of the world W, formalized as $R(x) \approx W$. Coherence concerns organizational compatibility: multiple interacting processes maintain sufficient mutual compatibility to preserve the integrity of a larger organizational structure, which we can express as $C(x_1, \ldots, x_n) > \epsilon$ for some threshold ϵ.

An organism can maintain coherence while sacrificing correspondence. Human beings routinely distort memories, rationalize contradictions, and selectively filter information in ways that preserve organizational stability despite reducing representational accuracy. This is not cognitive failure but organizational success: the organism prioritizes the maintenance of its coherent trajectory space over the accuracy of its internal descriptions of the world. Conversely, systems can maintain high representational accuracy while lacking coherence entirely. A language model may produce locally accurate continuations while lacking any persistent organizational structure binding those continuations into a stable purposive trajectory.

The distinction is especially important because many predictive-processing accounts implicitly assume that sufficient correspondence automatically generates coherence. The present framework rejects this assumption. Coherence is not reducible to representational accuracy because coherence concerns the stability of dynamical relationships rather than the correctness of symbolic descriptions. A collection of accurate but dynamically unintegrated representations does not constitute an organized system. Integration, not accuracy, is the primary organizational requirement.

It is worth acknowledging the degree of independence between the two. In ecologically normal environments, coherence and correspondence tend to co-vary: an organism that maintains coherent admissibility will generally track the features of its environment with sufficient accuracy. The distinction becomes critical at the boundaries, under stress, or in adversarial conditions. A cognitive system under extreme metabolic depletion, social isolation, or environmental disruption will sacrifice correspondence to maintain coherence, producing the systematic distortions characteristic of motivated reasoning, memory confabulation, and perceptual constancy effects. The claim is not that coherence and correspondence are empirically independent across all conditions, but that coherence is organizationally prior: it is the condition under which correspondence-tracking remains possible, not its equivalent or its product. The practical implication is that systems which optimize correspondence in ecologically normal training distributions may fail abruptly when the distributional conditions that allowed correspondence to incidentally produce coherence no longer hold.

4.3 Trajectory Collapse and Organizational Pathology

The admissibility framework allows pathological conditions to be reformulated in geometrical rather than purely inferential terms. Rather than treating dysfunction exclusively as erroneous inference or distorted prediction, dysfunction may be understood as the collapse or deformation of an organism's admissible trajectory space.

Definition 4.1. *A* trajectory collapse *occurs when the admissible subspace $\mathcal{A}(t)$ contracts such that the organism loses access to previously available organizational trajecto-*

ries necessary for viable persistence.

This reformulation provides a unified geometric interpretation of a wide range of pathological phenomena. In depression, the organism's future trajectory space contracts: actions that previously appeared viable lose their accessibility, and the organism becomes trapped within a shrinking basin of low-energy organizational states. The problem is not primarily that the depressed organism holds false beliefs about the world but that its topology of possibility has narrowed to the point where purposive action becomes geometrically unavailable. In addiction, local stabilization loops come to dominate higher-order coherence maintenance: the organism repeatedly returns to a narrow attractor basin that preserves short-term regulatory stability while progressively destroying long-term admissibility. The addictive attractor is locally stable and globally catastrophic. In obsessive cognition, recursive stabilization becomes overconstrained: the system sacrifices exploratory flexibility in exchange for excessive local predictability, narrowing its admissible space by excessive self-binding rather than by external pressure.

The same geometric analysis extends beyond individual organisms. In institutional rigidity, stabilization operators that once preserved coherence become excessively crystallized, preventing adaptation to changing environmental conditions and eventually destabilizing the larger system they were originally designed to preserve. In AI hallucination, the absence of genuine organizational constraints means that the system has no admissibility boundary to enforce: it generates continuations that are statistically plausible within its training distribution but organizationally incoherent with respect to any persistent imperative, because no such imperative exists for it to maintain.

These examples illustrate that organizational pathology is often geometrical rather than merely inferential. The problem is not simply that the organism "believes false things" but that its admissible trajectory topology has become distorted, contracted, or fragmented in ways that preclude viable persistence.

4.4 Temporal Layering and Recursive Persistence

The concept of persistence developed throughout this essay should not be understood as static self-identity. Biological and cognitive systems persist precisely through transformation; their continuity is recursive rather than inertial. This distinction requires a richer treatment of temporality than is typically present in optimization-based models.

Let $\mathcal{A}_k(t)$ denote admissibility structures operating at different temporal scales k. Fast-timescale admissibility governs immediate sensorimotor viability: the moment-to-moment coordination of bodily action with environmental structure that keeps the organism physically viable. Intermediate-timescale admissibility governs habits, routines, and social coordination: the organizational patterns that allow the organism to navigate its social and practical environment without continuous deliberation. Long-timescale admissibility governs identity, institutions, and civilizational continuity: the deep structural constraints that make certain life-trajectories coherent and others self-undermining across the span of years and generations.

Persistence emerges through the recursive coordination of these nested temporal layers, formalized by the compatibility condition

$$\mathcal{A}_{k+1}(t) \supseteq \mathcal{A}_k(t).$$

An organism remains coherent not by preserving identical states through time but by maintaining compatibility relations across multiple evolving temporal horizons simultaneously. This is why Bergson's insistence on the irreducibility of duration [4], Whitehead's process ontology [5], and Simondon's account of individuation as continuous transduction [12] are philosophically relevant here: each, in different ways, insists that temporal existence is not the successive occupation of instantaneous states but the ongoing constitution of a process through its own dynamical history.

This layered temporality helps explain why purely myopic optimization systems fail to exhibit robust agency. Systems optimized only for immediate prediction or short-term reward frequently destabilize the larger temporal structures upon which long-term coherence depends. The problem of intelligence is therefore not merely the minimization of some local prediction error $E(t)$ but the maintenance of coherence across recursively nested temporal admissibility structures. An agent that succeeds at the fast timescale while failing at the intermediate and long timescales is not a partial success but an organizational failure, because the temporal layers are not independent: fast-timescale decisions shape which intermediate-timescale trajectories remain admissible, and intermediate-timescale habits shape which long-timescale life-trajectories remain open.

Varela's account of temporal constitution in embodied mind [13] is relevant here as well. The organism does not merely exist in time; it constitutes its own temporal horizon through its ongoing activity. The admissibility structure $\mathcal{A}_k(t)$ is not a fixed constraint imposed from outside but an evolving product of the organism's own history of activity and response. Cognition is therefore temporally self-constituting in a sense that no prediction-error account captures, because prediction-error accounts treat time as the parameter along which inference occurs rather than as the medium in which the organism's organizational form is continuously generated.

4.5 Symbolic Structures as Topological Glue

The clipboard framework developed in the following chapter can be clarified through a preliminary topological interpretation of symbolic systems. A symbolic artifact — an essay, diagram, proof, institution, or ritual — does not merely preserve information. It preserves navigability across discontinuities.

Suppose a cognitive trajectory $\gamma : [t_0, t_1] \to X$ is interrupted by perturbation, distraction, or temporal separation. Without stabilization, the trajectory may become irrecoverable because the local organizational conditions that sustained it no longer exist. A symbolic stabilization operator S preserves a partial local structure by implementing a map $S : U_i \to U_j$, where $U_i, U_j \subset X$ are locally coherent neighborhoods within the trajectory manifold. The symbolic structure acts as a gluing operator between otherwise disconnected regions of cognitive space, allowing trajectories to be resumed despite discontinuities in time, attention, embodiment, or context.

This interpretation explains why symbolic systems scale civilization itself. Mathematics, writing, law, archives, rituals, and institutions collectively function as distributed topological glue preserving coherence across generations. Civilization is therefore not merely accumulated information but recursively stabilized navigability: the ongoing maintenance of a shared admissible trajectory space large enough to sustain collective organizational complexity across the disruptions of time.

4.6 Why Prediction Systems Simulate Agency

One final clarification concerns why predictive systems can appear agentive despite lacking organizational persistence. The appearance of agency in sophisticated prediction systems is not a mystery but a structural consequence of what those systems are trained on.

Sufficiently advanced continuation systems can locally approximate the outputs of genuinely agentive systems because the statistical structure of human symbolic behavior already contains compressed traces of agency. Human language encodes purposive structure, social constraint, temporal continuity, emotional regulation, and ecological adaptation. A sufficiently sophisticated continuation engine trained on this data therefore reproduces many surface features of agency without instantiating the organizational conditions that generated those features originally. The simulation may reproduce observable patterns with extraordinary accuracy while lacking the organizational constitution that generates those patterns in biological systems.

The distinction resembles the difference between simulating a hurricane numerically and being a hurricane physically. The simulation may reproduce observable patterns — pressure gradients, wind speeds, precipitation distributions — with extraordinary accuracy while lacking the thermodynamic organization constitutive of the phenomenon itself. The simulation has no eye, no latent heat release, no Coriolis-coupled convective dynamics: it has a numerical representation of the outcomes of those processes. Likewise, prediction systems can reproduce symbolic traces of agency without maintaining admissible trajectories, embodied persistence, or ecological coherence. Their apparent purposiveness is therefore derivative rather than intrinsic: it is borrowed from the organizational achievements of the biological agents whose symbolic output constitutes their training data, rather than generated by any organizational imperative of their own.

This observation has implications for AI alignment that deserve to be made explicit. Aligning a prediction system by training it to produce outputs that correspond to human preferences is not the same as creating a system that maintains a coherent relationship with human organizational imperatives over time. Correspondence-based alignment may succeed locally and fail globally, for the same reason that local predictive optimization can succeed while destroying global admissibility. An alignment framework adequate to the organizational account of agency developed here would need to stabilize a shared admissible field between human and system rather than optimize the correspondence between system outputs and human preference signals.

4.7 What This Framework Does and Does Not Claim

Four potential misreadings of the admissibility framework recur with enough frequency that it is worth addressing them directly before proceeding to the positive account of symbolic stabilization.

The first misreading is that the framework denies computation. The admissibility framework makes no such denial. Organisms perform operations that are computationally describable, exploit regularities, compress environmental structure, and stabilize local predictive loops. The issue is not whether computation occurs but whether computation constitutes the primary organizational principle of cognition. The admissibility framework argues that computation is subordinate rather than sovereign: computational operations emerge within an already existing persistence structure whose primary imperative is the maintenance of viable organizational trajectories. This can be expressed through the inclusion relation $\mathcal{C} \subset \mathcal{A}$, where $\mathcal{C}$ denotes the class of computationally tractable regulatory operations and $\mathcal{A}$ denotes the admissible organizational structure within which those operations remain viable. Computation appears not as the constitutive substrate of cognition but as one family of local stabilization mechanisms embedded within a larger persistence geometry. A neural subsystem may behave computationally without cognition itself being reducible to computation, in the same way that fluid turbulence may exhibit locally linear approximations while remaining globally nonlinear.

The second misreading is that persistence is a return to classical homeostasis. Homeostasis describes the maintenance of stable internal variables around equilibrium points. Persistence, in the sense developed here, is structurally richer. A homeostatic system attempts to preserve local equilibrium conditions; a persistence structure preserves the conditions under which coherent trajectories remain navigable across changing environments, scales, and organizational transformations. The organism does not merely maintain static variables but recursively reorganizes itself to preserve viability under conditions where equilibrium itself may require continual transformation. A developing organism cannot remain viable by preserving a fixed state; it must continuously alter its own organization while preserving coherence through the alteration. Formally, homeostasis seeks bounded deviation around a fixed attractor, $|x(t) - x^*| < \epsilon$, while persistence concerns the preservation of admissible navigability, $\gamma(t) \in \mathcal{A}(t)$, where both the trajectory and the admissible region evolve recursively. Persistence is therefore fundamentally processual rather than equilibrium-based.

The third misreading is that symbolic systems are epiphenomenal relative to biological organization. The stabilization-operator account rejects this hierarchy. Symbolic systems become organizationally real precisely insofar as they alter admissible trajectory structure. A legal institution changes which social trajectories remain viable. A scientific theory reorganizes which conceptual trajectories become accessible. A personal narrative alters which actions appear coherent to the organism maintaining it. Let $S : X \to X'$ be a symbolic stabilization operator acting over trajectory space. The symbolic structure becomes organizationally real when the induced admissible geometry satisfies $\mathcal{A}'(t) \neq \mathcal{A}(t)$: the symbolic operation has changed the topology of viable trajectories. Under this interpretation, symbolic systems are higher-order persistence structures capable of stabilizing trajectories across temporal and social scales inaccessible to purely biological regulation alone. Civilization becomes possible only because symbolic stabilization operators preserve coherence across generations that no individual organism could maintain independently.

The fourth misreading is that agency is a larger-scale optimization. The distinction between optimization and agency concerns irreversible organizational consequence, not merely scope or power. Optimization compresses trajectory space toward a minimum, $X \to x^*$; agency preserves navigable organizational possibility, $X \to \mathcal{A}(t)$. A sufficiently powerful optimizer may become increasingly dangerous precisely because it can discover trajectories exiting the shared admissible region while still satisfying its local objective function. Scaling optimization increases local search power without generating persistence structure. Genuine agency therefore requires categorical constraint maintenance rather than unrestricted scalar minimization, and this is a structural requirement independent of capability scale.

4.8 Operational and Conceptual Clarifications

Three further clarifications address concerns that the formalism raises but the preceding sections do not fully resolve.

Prediction as embedded rather than autonomous. The admissibility framework denies that prediction constitutes the constitutive ontology of cognition, but it does not treat prediction as a single monolithic operation to be accepted or rejected wholesale. When an organism generates a numerical estimate of when a ball will arrive, this is a locally defined predictive operation $P_i : U_i \subset X \to \mathbb{R}^{n_i}$ over a restricted neighborhood of trajectory space. This operation is not reducible to anticipatory entrainment, which is a matter of dynamical coupling rather than explicit estimation. The admissibility framework treats both as legitimate organizational operations that differ in their computational structure while sharing their organizational function: both serve the preservation of $\gamma(t) \in \mathcal{A}(t)$. The framework is therefore not a single-mechanism account of anticipation but a hierarchical account in which the global admissibility constraint is prior, and multiple distinct anticipatory mechanisms — entrainment, explicit estimation, habit, social coordination — are organized as subordinate strategies within it. When local prediction and admissibility maintenance conflict, admissibility takes precedence; when they are compatible, both operate. This interaction is not a gap in the framework but its normal operation.

Operational criteria for $\mathcal{A}(t)$. The admissible region $\mathcal{A}(t) \subset X$ should not be interpreted as a hidden object directly observable in isolation, any more than a dynamical attractor or a phase space manifold is directly observable. One reconstructs $\mathcal{A}(t)$ from the stability structure of observed trajectories under perturbation. Three observable signatures constrain its reconstruction. First, perturbation recovery structure: a system with stable admissibility returns to characteristic organizational regions after bounded perturbation, $\gamma(t+\delta) \to \mathcal{N}(x^*)$, with recovery dynamics that are not fully determined by any local loss landscape. Second, trajectory accessibility asymmetry: the framework predicts that some trajectories are dynamically inaccessible despite being locally optimal under restricted objective functions, producing the hysteresis and irreversibility characteristic of addiction, ecological collapse, and institutional rigidity. Third, stratified transition behavior: if $\mathcal{A}(t)$ possesses singular organizational boundaries, transitions between organizational modes should exhibit discontinuous accessibility changes rather than smooth parametric interpolation. These signatures are not sufficient to fully specify $\mathcal{A}(t)$, but they distinguish the admissibility account from a pure loss-landscape account in empirically testable ways. The framework is therefore not unfalsifiable; it is underdetermined in its current form and requires further formalization, which the appendices and the broader research program they gesture toward are designed to advance.

Organizational integrity as constraint closure. The concept of organizational integrity appears throughout this essay and requires explicit specification. It does not denote homeostatic stability, informational integration alone, or autopoietic closure in the restricted metabolic sense. Organizational integrity denotes recursive constraint closure across interacting organizational scales: the system's activity contributes to preserving the boundary conditions that allow the activity to continue occurring.

Definition 4.2. *A system possesses* organizational integrity *if the processes constituting the system recursively maintain the constraints necessary for the continued persistence of those same processes across time. Formally, let* $\mathcal{C}(t) = \{C_1, \ldots, C_n\}$ *denote the active constraint set governing a system's persistence. Organizational integrity exists when the dynamical evolution satisfies the recursive closure condition*

$$\forall\, C_i \in \mathcal{C}(t), \quad \exists\, \gamma_j \text{ such that } \gamma_j \to C_i.$$

That is, the system's trajectories regenerate the constraints sustaining those trajectories themselves.

This distinguishes persistence structures from optimization systems. A predictive model may maintain local statistical coherence while lacking recursive constraint closure: its outputs need not preserve the organizational conditions enabling future output generation. Biological organisms continuously regenerate metabolic, ecological, behavioral, and symbolic boundary conditions necessary for their own persistence. Organizational integrity is therefore not a vague explanatory primitive but a precisely characterizable structural property: it is present when recursive constraint closure holds, and absent when it fails. The concept is related to but distinct from autopoiesis, which concerns metabolic self-production; organizational integrity as defined here extends to symbolic, institutional, and social constraint structures that metabolic accounts do not reach.

Chapter 5

The Clipboard as a Stabilization Operator

5.1 Memory Beyond Storage

The standard account of memory in cognitive science treats it as a storage-and-retrieval system: information is encoded in some substrate, stored across time, and retrieved when needed. This account is deeply shaped by the computer metaphor: memory is RAM, long-term memory is disk storage, retrieval is read-access. Even more sophisticated accounts — connectionist, embodied, or enactivist — tend to preserve the basic functional architecture of storage and retrieval while relocating it from explicit data structures to distributed activation patterns or bodily habits.

This essay proposes a different account, which we can call the clipboard thesis. The central claim is that what cognitive systems call "memory" is better understood as a family of stabilization operators: structures that preserve coherence across temporal gaps, anchor organizational continuity in the face of perturbation, and maintain admissible trajectories through otherwise volatile transformations. A clipboard, in the broadest sense, is any structure that performs this function — any device that keeps a trajectory open when the forces that generated it have passed and the forces that will continue it have not yet engaged.

The familiar digital clipboard is a useful but misleading exemplar. What the clipboard does in a computational system is preserve the content of a selection across the temporal gap between copy and paste. But this preservation is achieved by storing a static representation of the content, which is retrieved unchanged on demand. The cognitive phenomenon that the clipboard metaphor is meant to capture is not like this. What is preserved in genuine cognitive continuity is not static content but organizational coherence: the capacity to continue a trajectory that was interrupted, to pick up a line of thought after a distraction, to return to a project after sleep, to maintain purposive direction across the discontinuities of daily life.

This kind of continuity is not achieved by storing representations. It is achieved by maintaining stabilizing structures that make certain trajectories available and others unavailable: notes, sketches, marked pages, organized desks, familiar environments, ritual practices, conversational partners, and institutional frameworks. These are not memory stores in the computational sense; they are organizational scaffolds that anchor trajectories by distributing their stabilization across the organism's environment.

Definition 5.1. *A* stabilization operator *is any structure — internal, external, social, or symbolic — that preserves the availability of an organizational trajectory across a temporal gap during which the trajectory cannot be actively maintained by the organism alone.*

The physical substrate of a stabilization operator varies by scale and organizational level, and this variation is not an embarrassment but a consequence of the framework's generality. At the neural level, synaptic potentiation is a stabilization operator whose physical substrate is the changed conductance of synaptic channels. At the cognitive level, a written note is a stabilization operator whose physical substrate is ink on paper, pixels on a screen, or any durable encoding of constraint structure. At the institutional level, a constitution is a stabilization operator whose physical substrate is a combination of written documents, enforcement mechanisms, social practices, and the trained dispositions of institutional actors. What unifies these across their substrate diversity is the functional property: each preserves the availability of an organizational trajectory across a temporal gap by externalizing enough of its constraint structure to allow re-entry. The appropriate question is therefore not what single physical substance constitutes a stabilization operator, but what physical conditions are sufficient to preserve constraint structure across the relevant temporal gap — a question that admits different answers at different organizational scales without undermining the framework's unity.

5.2 Essays, Notes, and Recursive Constraint Stabilization

The symbolic structures that cognitive systems routinely produce — essays, notes, diagrams, programs, mathematical proofs, musical scores — function as stabilization operators in a specific and important way: they preserve the constraint structure of a trajectory rather than its content. An essay does not store the thought it expresses; it re-creates the organizational conditions under which the thought can be re-inhabited. Reading an essay by a thinker is not retrieving stored information; it is entering into the constraint-field that the thinker has articulated, and using that constraint-field to navigate toward the organizational state in which the thinker's trajectory becomes available.

This can be formalized as follows. Let X be the cognitive trajectory space of an organism, and let $\mathcal{A}(t) \subset X$ be the organism's admissible subspace at time t. An essay E is a symbolic structure that implements a compression mapping

$$f_E : X \to M_E,$$

where M_E is a compressed representation of the constraint structure of a specific region of X. The essay does not preserve trajectories directly; it preserves enough of the constraint structure of a region of X to make those trajectories re-accessible to a reader who enters the field generated by f_E. Reading the essay is the inverse operation f_E^{-1}, which expands the compressed constraint structure back into a navigable region of the reader's trajectory space.

This account explains why good essays are not merely informative but cognitively generative: they do not just convey propositions but reorganize the reader's admissible trajectory space in ways that make new organizational states accessible. They function as constraint-maps rather than content-stores. The value of a good essay is not that it tells the reader something they did not know but that it reconfigures the topology of the reader's cognitive space in ways that make new trajectories available.

The same account applies to mathematical proofs, which are perhaps the clearest case of symbolic constraint stabilization. A proof is not a record of the mathematician's mental states; it is a sequence of constraint-preserving transformations that takes the reader from the axioms to the theorem by a path that is organizationally stable enough to be traversed repeatedly and shared across different cognitive systems. The proof preserves the constraint structure of the mathematical trajectory from hypothesis to conclusion, making that trajectory available to any reader who enters the field defined by the proof's axioms and rules of inference.

Proposition 5.2. *Symbolic structures that function as stabilization operators implement recursive constraint stabilization: they not only preserve specific trajectories but preserve the organizational conditions under which new trajectories can be generated.*

Proof. Let E be a symbolic structure implementing a compression mapping $f_E : X \to M_E$. Suppose E preserves the constraint structure of a region $R \subset X$. A reader who enters the field of E gains access not only to the trajectories explicitly encoded in M_E but to the constraint structure of R itself, which determines the topology of trajectories in R. This topology includes not only the specific trajectories that E represents but all trajectories that are consistent with the constraint structure of R. The reader can therefore generate new trajectories within R by navigating the constraint structure that E has made available, rather than merely retrieving the specific trajectories that E explicitly encodes. Symbolic structures that function as stabilization operators thus implement recursive constraint stabilization: they preserve not only trajectories but the generative conditions for new trajectories. $\square$

Rituals, institutions, and identities function analogously. A ritual preserves the organizational conditions under which a collective state can be re-inhabited; it is a recursive stabilization operator for a shared admissible trajectory space. An institution encodes constraint structures that make certain social trajectories stable and others unstable; it is a topological anchor for a collective organizational field. An identity is the persistence of a characteristic constraint structure through the disruptions of time and circumstance; it is a recursive self-stabilization operation that makes certain trajectories consistently available and others consistently unavailable.

5.3 The Geometry of Selfhood

The account of memory and symbolic structure developed in the preceding sections leads naturally to a reconceptualization of selfhood. If stabilization operators are the primary cognitive structures, then the self is not a stored representation of a persistent entity but the persistence of a characteristic constraint structure through time.

Definition 5.3. Selfhood *is the persistence of a coherent constraint structure across temporal gaps and organizational perturbations: the maintenance of a characteristic admissible trajectory space whose topology is stable enough to be recognizable as continuous, while dynamic enough to accommodate growth, change, and response to circumstance.*

On this account, self-identity is not a metaphysical puzzle about the persistence of a substance through time but a dynamical fact about the topology of an admissible trajectory space. The self persists insofar as the constraint structure of its admissible trajectory space remains coherent: insofar as certain trajectories remain available and certain others remain unavailable across the disruptions of sleep, distraction, illness, and change. The self is disrupted insofar as this constraint structure becomes incoherent: insofar as trajectories that were previously available become closed, or previously closed trajectories become open in ways that destabilize the organism's organizational coherence.

This account preserves the intuition that personal identity is connected to continuity of experience, character, and purpose, while replacing the problematic notion of a persistent mental substance with the more tractable notion of a persistent constraint topology. The self is not a thing but a dynamical pattern: a stable mode of navigating an admissible trajectory space that persists through time by continuously re-instantiating the conditions of its own coherence.

Narrative plays a crucial role in this self-stabilization process. The stories an organism tells about itself are not accurate records of what happened but constraint-maps that make certain trajectories available and others unavailable. A person who understands themselves as a scientist, an artist, or a survivor has a self-narrative that imposes a constraint structure on their admissible trajectory space, making certain actions coherent and others incoherent given who they understand themselves to be. The self-narrative is a stabilization operator for the self's characteristic constraint topology.

Chapter 6

Toward a Topological Ontology of Cognition

6.1 Preliminary Clarifications

Before developing the topological ontology in formal detail, it is worth addressing directly a cluster of objections that the preceding chapters invite but have not yet fully answered. These concern the explanatory status of geometrical language, the distinction between descriptive and constitutive accounts, the relationship between local and global viability, and the sense in which topology rather than symbolic computation is the appropriate framework for cognition. Clearing these matters explicitly prevents the more formal sections from being read as decorative rather than structural.

Why the Framework is Not Merely Metaphorical

One possible objection is that the topological language employed throughout this essay is merely metaphorical: that terms such as trajectory space, boundary, stratification, and organizational collapse redescribe familiar psychological phenomena without introducing genuinely explanatory structure. This objection misunderstands the role that geometry already plays in contemporary scientific explanation.

Modern physics characterizes systems geometrically rather than mechanistically as a matter of course. General relativity replaces gravitational force with spacetime curvature. Dynamical systems theory characterizes organization through attractors, bifurcations, and phase transitions rather than through symbolic rules.

Control theory defines viability kernels geometrically as constrained regions of state-space within which persistence remains possible. The admissibility framework extends this style of explanation into cognition and agency, and in doing so it stands on established methodological ground rather than novel metaphorical ground.

The admissible region $\mathcal{A}(t) \subset X$ is not a metaphor for viability; it is the organizational condition under which viability exists. Trajectory collapse is not an evocative phrase for pathology but a claim about the contraction of accessible organizational states. Depression, addiction, rigidity, and institutional failure all exhibit the same formal structure: a narrowing of accessible trajectories and a reduction in the dimensionality of viable organizational movement. The framework is therefore an attempt to identify the geometrical substrate underlying persistence itself, not a symbolic redescription of substrate already well characterized by other means.

The Distinction Between Description and Constitution

A recurrent confusion throughout computational neuroscience arises from the failure to distinguish between descriptive adequacy and constitutive ontology. A model may successfully describe the behavior of a system without constituting the causal architecture by which the system operates. Predictive-processing frameworks frequently move illegitimately from the claim that M predicts X to the claim that X internally implements M. The success of Bayesian or predictive models becomes reinterpreted as evidence that organisms literally perform Bayesian inference internally. But this inference does not follow.

The weather is accurately modeled through Fourier decomposition without the atmosphere itself containing Fourier coefficients. Orbital mechanics is modeled through coordinate systems that planets do not instantiate. Likewise, predictive-processing models may successfully characterize statistical regularities in neural activity without the nervous system implementing predictive coding as its constitutive ontology. The scientist's compressed model $M = \pi(X)$ remains a projection of the organism's ecological trajectory space rather than a proof that the organism computes over M internally.

This distinction is not merely philosophical but methodological. Entire research programs become distorted when descriptive tools are mistaken for constitutive architecture, because scientists begin searching for neural implementations of operations that exist only in the observer's analytical framework. The reification error documented in Chapter 1 is not an occasional mistake but a systematic consequence of ignoring the description/constitution distinction.

Why Local Success Does Not Imply Global Viability

One of the central organizational principles of this essay is that local optimization and global persistence are fundamentally different kinds of achievement. Biological systems frequently tolerate local inefficiency to preserve larger-scale organizational coherence, while optimization systems often destroy their own viability precisely by succeeding too effectively at local objectives. An addict locally stabilizes affective distress while globally collapsing long-term viability. A bureaucracy optimizes measurable metrics while destroying the institutional flexibility necessary for survival. A language model maximizes statistical continuation accuracy without maintaining any coherent organizational trajectory beyond the continuation itself.

The mathematical distinction is already established in the preceding chapter: local optimization seeks the best point in an unrestricted space, $\gamma^* = \operatorname{argmin}_{\gamma \in X} E(\gamma)$, while admissibility requires preservation of movement within a constrained organizational region, $\gamma(t) \in \mathcal{A}(t)$. Optimization pressure frequently drives systems toward trajectories that are locally efficient but globally catastrophic, which is precisely why local inferential success cannot serve as a sufficient criterion for cognition. Cognition is not the production of accurate local outputs but the preservation of coherent admissible trajectories across time.

Topology Rather Than Symbol Manipulation

The deeper implication of the admissibility framework is that cognition is more naturally understood as topological navigation than as symbolic computation. Symbolic systems presuppose stable external semantics: symbols function only insofar as their mappings remain coherent across transformations. Biological cognition must maintain the conditions under which such coherence remains possible in the first place. An organism does not merely manipulate representations of the world; it maintains the viability of the interface through which a world can continue appearing as navigable at all.

Stability basins, attractor structures, stratified boundaries, organizational folds, and admissible transitions are not secondary descriptive conveniences layered atop symbolic cognition; they are the deeper geometrical conditions under which symbolic activity becomes possible. This is why organisms can survive persistent representational inaccuracy while remaining organizationally coherent, whereas purely predictive systems may exhibit extraordinary representational competence while lacking persistence entirely. Representation is therefore subordinate to organization:

$$\text{Representation} \subset \text{Organizational Persistence}.$$

The central problem of cognition is not how symbols correspond to the world but how organized systems remain coherent long enough for meaningful symbolic relations to emerge at all.

6.2 From Logic to Topology

The dominant mode of cognition in the Western philosophical and scientific tradition is logical: it takes the fundamental cognitive operation to be the manipulation of truth-functional representations according to rules of inference. On this account, cognition is primarily a matter of constructing, evaluating, and acting on propositions about the world. The computational metaphor is the contemporary expression of this logical tradition: it inherits logic's commitment to discrete, context-independent, compositionally structured representations, and adds the claim that the brain is the physical implementation of a logical processor.

This essay proposes an alternative: that cognition is fundamentally topological rather than logical. The fundamental cognitive operation is not the evaluation of truth-functional propositions but the management of coherence-preserving deformations within an admissible trajectory space. The cognitive system is not primarily a truth-machine but a coherence-preservation machine: it is organized to maintain the topological integrity of its organizational structure through a dynamically changing environment.

The difference between logical and topological frameworks can be illustrated by considering their different accounts of equivalence. In a logical framework, two cognitive states are equivalent if they have the same truth-functional content: if they represent the same propositions as true. In a topological framework, two cognitive states are equivalent if they are continuously deformable into each other within the admissible trajectory space: if there is a path from one to the other that does not cross the boundary of the admissible region. This topological notion of equivalence is much richer than the logical one: it is sensitive to the shape of the admissible region, to the trajectories that connect different states, and to the organizational transformations that a cognitive system can perform.

The shift from logical to topological ontology is not merely a change of mathematical formalism. It reflects a deep reorientation in what cognition is understood to be about. Logical cognition is about truth: the cognitive system is organized to track what is the case, to infer what follows from what, and to act on accurate representations of the world. Topological cognition is about coherence: the cognitive system is organized to maintain the structural integrity of its own organization, to navigate admissible trajectories, and to preserve the conditions of its own continued functioning.

These are not mutually exclusive imperatives. A cognitive system that successfully maintains its organizational coherence will, as a consequence, tend to track truth in the relevant domains. But the relationship is indirect: truth-tracking is a consequence of coherence maintenance, not the primary imperative. And when the two come into conflict — when maintaining organizational coherence requires ignoring or distorting truth — biological systems consistently prioritize coherence. The pervasiveness of cognitive biases, motivated reasoning, and confabulation in human cognition is not an anomaly to be explained away; it is evidence that the primary imperative of human cognition is coherence maintenance rather than truth-tracking.

6.3 Manifolds of Agency

Agency, on the topological account, is the capacity to maintain coherent admissible trajectories under changing constraints. This definition has several important properties that distinguish it from standard accounts of agency in philosophy and cognitive science.

Definition 6.1. Agency *is the capacity of a system to maintain the coherence of its admissible trajectory space across environmental perturbations: to sustain a characteristic organizational structure that makes certain trajectories available and others unavailable, and to navigate that structure in ways that preserve organizational integrity.*

This definition makes agency a matter of degree rather than a binary property. Systems differ in the richness of their admissible trajectory space, in the sophistication of their capacity to navigate that space, and in the robustness of their organizational coherence under perturbation. A bacterium has a relatively simple admissible trajectory space and a limited capacity to navigate it, but it is genuinely agentive: it maintains its organizational coherence through dynamical coupling with its environment and actively manages the conditions of its own continued existence. A human being has an enormously rich admissible trajectory space, sophisticated navigation capacities, and a highly robust organizational structure, but these are differences of degree, not kind.

The topology of the admissible trajectory space determines the structure of agency. Key topological features include stability basins: regions of the trajectory space from which the system reliably returns to a characteristic organizational state after perturbation. A stability basin is not merely a set of states; it is a dynamical structure characterized by the existence of a stable attractor at its center and a boundary beyond which the system is drawn toward a different attractor or organizational collapse. Cognitive systems maintain multiple stability basins simultaneously, corresponding to different organizational modes, and navigate between them in ways that preserve overall organizational coherence.

Another key topological feature is the organizational fold: a region of the trajectory space where nearby trajectories diverge sharply, corresponding to a critical choice or transition point in the organism's organizational history. At an organizational fold, small differences in the organism's state at time t lead to large differences in its organizational trajectory over subsequent times. The organism must navigate through folds carefully, because errors at a fold can lead to trajectories that exit the admissible subspace entirely.

Dynamical coherence structures are perhaps the most important topological feature. A coherence structure is a persistent pattern of mutual constraint among the components of the organism's organizational system that maintains the integrity of the whole across the perturbations of the parts. The immune system, for example, is a dynamical coherence structure: it maintains the integrity of the organism's biological identity by continuously discriminating self from non-self and responding to threats to organizational integrity. Cognitive systems have analogous coherence structures at multiple scales, from the neural mechanisms that maintain the coherence of perceptual experience to the social and cultural mechanisms that maintain the coherence of collective identity.

6.4 Cognition as Boundary Management

The topological account of cognition makes boundary management the central cognitive operation. The organism-environment boundary is not a fixed line but a dynamically evolving interface that the organism must actively maintain in order to preserve its organizational integrity. Cognition, on this account, is the ongoing management of this interface: the continuous regulation of what crosses the boundary, in what form, with what transformation, and with what consequences for the organism's organizational structure.

Definition 6.2. *The* organism-environment interface *at time t is the dynamically evolving boundary $\partial O(t)$ between the organism's organizational space and its environment, characterized by the flows of matter, energy, information, and constraint that cross it and by the transformations the organism imposes on those flows.*

Affordances, in Gibson's sense, are features of the organism-environment interface: they are properties of the interface that make certain actions available and others unavailable given the organism's specific bodily organization and action-capacities. An affordance is not a property of the environment alone, nor a property of the organism alone, but a relational property of the organism-environment system at the interface. This is why affordance perception is not inference: it is the organism's direct registration of its own interface topology, of what the environment affords given what the organism can do.

Embodiment, on this account, is not merely the fact that cognitive systems have bodies but the fact that the organism's bodily organization is the primary determinant of its interface topology. Different bodies make different affordances available; the organism's morphology, physiology, and action-capacities shape the topology of its admissible trajectory space in ways that are not separable from the organism's cognitive organization. Cognition is not a process that happens inside the organism and then interacts with the body; it is a process constituted by the organism's ongoing management of its bodily interface with the environment.

Ecological intelligence, on this account, is not abstract computational power but the capacity for sophisticated interface management: the ability to maintain a rich and coherent admissible trajectory space through a dynamically changing environment by means of flexible, context-sensitive management of the organism-environment interface. An ecologically intelligent organism is not one that has the most accurate world-model but one that can navigate the widest range of environments while maintaining organizational coherence.

6.5 Admissibility, Symmetry, and the Local Linearization of Organization

The admissibility framework developed throughout this essay implicitly relies on a distinction that also appears in modern geometry and mathematical physics: the distinction between a globally curved organizational structure and the locally linear approximations through which that structure becomes analytically tractable. The relationship between Lie groups and their associated Lie algebras provides a precise formal analogy for clarifying this distinction, and ultimately for clarifying why local predictive operations cannot, in principle, fully characterize global organizational persistence.

A Lie group is simultaneously a group and a smooth manifold. It possesses both algebraic and continuous geometrical structure: rotations, continuous symmetries, gauge transformations, and smooth dynamical flows are all naturally described in this language. Lie groups are, however, frequently too globally complex to study directly. One approach is to study instead the tangent structure at the identity element, T_eG, which forms the associated Lie algebra $\mathfrak{g}$. The Lie algebra acts as a local linearization of the larger nonlinear structure: one studies the curved global object indirectly through its infinitesimal generators, the local directional flows that collectively determine the global geometry.

This distinction is philosophically important for cognition because predictive-processing frameworks implicitly assume that local inferential operators fully characterize the global organization of the system. The admissibility framework developed here instead proposes that local predictive operations are tangent approximations to a larger organizational manifold whose global structure cannot be recovered from the properties of its local linearizations alone. A local prediction operator behaves analogously to an infinitesimal tangent approximation:

$$P_t : T_xX \to T_{x+\delta}X.$$

It estimates immediate local evolution within a restricted neighborhood of the trajectory manifold. Admissibility, by contrast, concerns the topology of the manifold itself, $\mathcal{A}(t) \subset X$, including its folds, singularities, attractors, bifurcations, and coherence boundaries. Local prediction may approximate movement along the manifold without determining the manifold's global organizational structure.

This distinction clarifies why organisms can remain globally coherent despite persistent local predictive failure. Biological systems frequently tolerate noisy, inaccurate, or incomplete local predictions while preserving larger-scale organizational integrity. Conversely, systems may achieve excellent local prediction while globally destabilizing themselves — a condition increasingly characteristic of optimization architectures whose local inferential success gradually destroys the larger admissibility structures upon which their continued viability depends. Just as knowing the tangent space at every point of a Lie group does not by itself determine the global topology of the group, knowing the local predictive dynamics of a cognitive system does not determine its global organizational persistence structure.

6.6 The $\mathfrak{sl}_2$ Seed Structure and Organizational Modes

The simplest nontrivial semisimple Lie algebra, $\mathfrak{sl}_2$, provides a useful conceptual model for understanding recursive organizational structure because it generates stable transformational hierarchies through only three generators [14]:

$$H, \quad E, \quad F,$$

satisfying the bracket relations

$$[H, E] = 2E, \qquad [H, F] = -2F, \qquad [E, F] = H.$$

Here H measures organizational state, while E and F generate upward and downward transformations through the state hierarchy. In the representation theory of $\mathfrak{sl}_2$, finite-dimensional modules decompose into weight spaces V_λ indexed by eigenvalues of H, with the actions of E and F moving vectors between neighboring weight spaces:

$$E : V_\lambda \to V_{\lambda+2}, \qquad F : V_\lambda \to V_{\lambda-2}.$$

The weight-space decomposition $V = \bigoplus_\lambda V_\lambda$ induces a partially ordered organizational geometry — a navigable transformational manifold structured by admissible transitions between neighboring organizational regions:

$$\cdots \leftrightarrow V_{\lambda-2} \leftrightarrow V_\lambda \leftrightarrow V_{\lambda+2} \leftrightarrow \cdots$$

The analogy to cognition is not that cognitive systems literally implement $\mathfrak{sl}_2$ representations, but that organizational systems frequently exhibit recursively generated admissibility hierarchies whose local transformational operators determine the accessibility relations between global organizational states. Certain cognitive transformations increase abstraction, energetic flexibility, or exploratory openness; others consolidate or stabilize organizational states into narrower attractor basins. Sequences of thought often proceed through neighboring admissible transformations rather than arbitrary jumps across cognitive space.

The highest-weight structure of finite-dimensional $\mathfrak{sl}_2$ representations is particularly suggestive. Every finite irreducible representation contains a highest-weight vector v_0 satisfying $E \cdot v_0 = 0$: an extremal coherence state from which the entire representation unfolds recursively through repeated applications of F,

$$v_k = F^k v_0,$$

generating the constrained organizational trajectory $v_0 \to v_1 \to \cdots \to v_n$. The finite-dimensionality condition forces organizational closure: since the representation cannot extend indefinitely, repeated applications of F must eventually terminate at $F \cdot v_n = 0$. The representation is therefore not an unrestricted expansion but a constrained navigable structure whose coherence depends on both generative openness and eventual closure.

Biological and cognitive systems exhibit analogous extremal structures. A developmental program begins from a tightly constrained organizational seed and recursively unfolds increasingly differentiated trajectories. A scientific paradigm begins from a compact generative framework whose recursive transformations produce extended conceptual families. A symbolic narrative similarly acts as a high-coherence organizational seed from which multiple subordinate interpretations and trajectories become available. The broader organizational principle that the $\mathfrak{sl}_2$ case illustrates is that viable structure is constrained generative navigability: neither pure rigidity, which destroys adaptability, nor pure openness, which destroys coherence, but bounded exploratory manifolds structured strongly enough to preserve identity while permitting adaptation.

It is worth being explicit about what this analogy does and does not constrain empirically. The $\mathfrak{sl}_2$ structure does not predict specific neural mechanisms or generate precise quantitative claims about firing rates or connectivity. What it does constrain is the organizational architecture of admissibility hierarchies: any system that exhibits bounded generative navigability with raising and lowering transitions between qualitatively distinct organizational strata will be subject to the same topological closure conditions that make finite-dimensional representations finite-dimensional. Empirically, this generates the prediction that cognitive systems undergoing development or therapeutic reorganization should exhibit discrete stratum transitions rather than continuous parametric improvements, that there exist ceiling states from which exploratory operators produce no further differentiation, and that descent from higher to lower strata should be dynamically easier than ascent. These predictions are testable, though their formalization requires the stratified admissibility framework developed in the following section rather than the Lie algebra alone.

6.7 Polynomial Realizations and the Geometry of Transformability

Every finite-dimensional irreducible $\mathfrak{sl}_2$ representation can be realized concretely through homogeneous polynomials in two variables:

$$\{x^a y^b : a + b = n\},$$

where the operators act as differential operators on this polynomial space:

$$H = x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y}, \qquad E = x\frac{\partial}{\partial y}, \qquad F = y\frac{\partial}{\partial x}.$$

The action of E raises the power of x while lowering the power of y, and conversely for F:

$$E(x^a y^b) = b\,x^{a+1}y^{b-1}, \qquad F(x^a y^b) = a\,x^{a-1}y^{b+1}.$$

The representation therefore behaves as a constrained flow across a polynomial manifold whose coordinates continuously transform under the admissibility rules generated by the algebra.

This realization is philosophically important because it reveals that the representation is fundamentally geometric rather than merely symbolic. The generators do not manipulate abstract labels; they deform structured spaces through continuous transformational flows. The significance for cognition is that symbolic reasoning itself often behaves less like discrete rule execution and more like constrained deformation within a structured semantic field. Concepts transform into neighboring concepts through locally admissible operations while preserving larger coherence relations; mathematical reasoning, metaphorical transfer, narrative development, and scientific abstraction all exhibit this partially geometric structure.

The polynomial realization also demonstrates why local transformations alone are insufficient to characterize global organization. The differential operators act locally on individual monomials, but the representation exists as a globally constrained transformational space whose structure emerges from the compatibility relations among all local operations simultaneously. A system may possess sophisticated local update rules while lacking globally coherent persistence structure: its outputs may be locally coherent while lacking the deeper persistence relations that bind biological trajectories into temporally unified organizational structures. The broader conclusion is therefore fully consistent with the central thesis of this essay:

$$\text{local inferential structure} \neq \text{global organizational coherence}.$$

Lie theory becomes philosophically important here precisely because it formalizes how globally coherent structures emerge from recursively constrained local transformations. Understanding the local generators alone is not enough. One must understand the topology of the larger admissible manifold whose persistence those generators collectively sustain.

6.8 Stratified Admissibility and Singular Organizational Boundaries

The treatment of $\mathcal{A}(t)$ as a smooth submanifold of X is an idealization adequate for much of the preceding analysis but insufficient for a full account of organizational pathology, attractor collapse, and identity transitions. In the generic biological case, the admissible trajectory space is not smoothly connected. It contains singular boundaries, discontinuous phase transitions, and regions of qualitatively different organizational character separated by boundaries across which trajectories cannot pass continuously. The appropriate mathematical structure is therefore not a smooth manifold but a stratified space.

Definition 6.3. *The* stratified admissibility decomposition *of the trajectory space X*

is a partition

$$\mathcal{A}(t) = \bigcup_i \mathcal{S}_i(t),$$

where each stratum $\mathcal{S}_i(t)$ is a smooth submanifold of X and the closure of each stratum is contained in the union of strata of equal or lower dimension. Strata are ordered by the organizational coherence they support: higher strata correspond to richer organizational modes and lower strata to more degenerate or constrained organizational conditions.

The stratification structure captures several phenomena that a smooth admissibility manifold cannot adequately represent. First, it formalizes the existence of qualitatively distinct organizational modes within a single organism: waking and sleeping, focused and diffuse, stressed and regulated, creative and rigid. These are not merely points on a continuum but distinct strata separated by transition boundaries. An organism navigating from one stratum to another undergoes a qualitative organizational transition that is not continuously deformable within a single stratum.

A legitimate methodological concern is how stratum boundaries are detected rather than stipulated post hoc. The framework proposes three empirical criteria for distinguishing genuine stratum boundaries from steep but continuous gradients. First, transitions across genuine stratum boundaries should exhibit hysteresis: the system follows different paths during descent and ascent, so that the conditions required to exit a lower stratum differ from the conditions that led to entering it. This asymmetry is absent from continuous gradient dynamics where the same path can be traversed in both directions. Second, genuine stratum boundaries produce discontinuous changes in trajectory accessibility: qualitatively different actions or cognitive operations become available or unavailable after transition, not merely more or less probable. Third, recovery from genuine stratum descent requires coordinated multi-constraint reorganization rather than parametric adjustment of a single variable; if fine-tuning a single parameter restores the prior organizational mode, the transition was a steep gradient, not a stratum crossing. These criteria are operational rather than definitional, and they admit degrees of evidence rather than yielding binary verdicts, but they give the stratification hypothesis empirical purchase distinguishable from mere redescription.

Second, stratification captures the asymmetry of organizational transitions. Moving from a higher stratum to a lower one — from a richer organizational mode to a more impoverished one — may be dynamically easy: perturbation, stress, resource depletion, or injury can all push an organism across a stratum boundary toward a lower stratum. Moving in the reverse direction may require sustained effort, external scaffolding, or favorable environmental conditions that reorganize the boundary structure of $\mathcal{A}(t)$ itself. This asymmetry corresponds to the empirical fact that degradation of organizational coherence is typically faster and easier than its restoration.

Third, the singular boundaries between strata are precisely where organizational pathology concentrates. Depression, addiction, psychosis, and acute stress responses are best understood not as displacements within a stratum but as transitions to lower strata — often accompanied by a collapse of the higher stratum's boundary structure that makes return difficult without external reorganization. The attractor basin structure visible within each stratum becomes inaccessible when the organism has crossed a stratum boundary downward, because the attractors of higher strata may not exist as stable configurations within lower strata.

Proposition 6.4. *In the stratified admissibility decomposition, organizational pathology is characterized by involuntary descent through stratum boundaries, and recovery requires the reconstruction of boundary conditions sufficient to support re-entry into higher strata.*

Proof. Let $\mathcal{S}_h(t)$ and $\mathcal{S}_l(t)$ be high and low strata respectively, with $\dim \mathcal{S}_h > \dim \mathcal{S}_l$. A trajectory γ occupying $\mathcal{S}_h$ undergoes pathological descent when perturbation forces γ across the boundary $\partial\mathcal{S}_h \cap \mathcal{S}_l$. Within $\mathcal{S}_l$, the dynamical structure is impoverished: the attractor landscape of $\mathcal{S}_h$ is generically unavailable because the boundary conditions that sustained those attractors are no longer satisfied. Recovery requires not merely local error correction — movement within $\mathcal{S}_l$ toward its own attractors — but the reconstruction of the boundary structure $\partial\mathcal{S}_h$, which depends on conditions external to the trajectory's current position in $\mathcal{S}_l$. This reconstruction is generically harder than descent because it requires coordinated reorganization of multiple coupled constraint relations rather than local gradient descent on a loss surface. $\square$

The stratification framework also clarifies the structure of identity transitions and developmental reorganization. Growth, maturation, creative breakthrough, and therapeutic transformation are all, on this account, transitions to higher or structurally richer strata: reorganizations that make new attractors, new trajectory families, and new organizational modes available that were inaccessible from the prior stratum. These transitions are not merely quantitative improvements but qualitative restructurings of the admissibility geometry. They require not just movement within an existing stratum but alteration of the stratum boundary structure, which is precisely what developmental programs, intensive learning, and transformative experience accomplish. The stratified decomposition therefore provides a geometric account of growth that is genuinely structural rather than merely descriptive: development is the progressive complexification of the organism's admissibility stratification.

The proposition's claim that recovery requires reconstruction of boundary conditions deserves elaboration, since the essay would otherwise be silent on the specific mechanisms by which such reconstruction occurs. Three classes of mechanism are relevant. The first is external scaffolding: the introduction of stabilization operators from outside the organism's current stratum that temporarily maintain boundary conditions the organism cannot sustain internally. Therapeutic relationships, institutional support, carefully structured environments, and ritualized practices all function in this way — they hold boundary structure open long enough for the organism's own organizational dynamics to re-stabilize within the higher stratum. The second is constraint reduction: the temporary relaxation of competing constraint pressures that have been consuming organizational resources needed for stratum maintenance. Rest, regulated nutrition, removal from stressful environments, and the reduction of competing demands all operate by freeing organizational capacity for boundary reconstruction. The third is trajectory seeding: the introduction of a constrained generative seed — a novel practice, a compelling narrative, a structured relationship, a disciplinary framework — that acts as a high-coherence anchor around which the higher stratum's attractor landscape can begin to reconstitute. Each of these mechanisms operates not by correcting local predictive errors within the lower stratum but by reconstructing the global boundary conditions under which the higher stratum becomes dynamically accessible. This is the precise sense in which recovery is not error correction but organizational restructuring.

The connection to the Lie-theoretic discussion of the preceding sections is direct. The weight-space decomposition $V = \bigoplus_\lambda V_\lambda$ of an $\mathfrak{sl}_2$ representation is itself a stratification: each weight space V_λ is a stratum, and the operators E and F generate admissible transitions between adjacent strata. The highest-weight condition $E \cdot v_0 = 0$ and the annihilation condition $F \cdot v_n = 0$ mark the singular boundaries of the representation: the strata beyond which no admissible transition exists. The finite-dimensionality of the representation corresponds, in organizational terms, to the existence of both a ceiling and a floor on organizational complexity — a bounded stratification within which the organism navigates.

6.9 Optimization, Admissibility, and the Structure of Organizational Imperatives

The admissibility framework's critique of prediction-error minimization can be sharpened by making explicit what distinguishes admissibility-based organization from optimization-based organization. This distinction is precise and carries significant consequences for the theory of cognition and the design of artificial systems.

Optimization in the relevant sense is the operation

$$\gamma^* = \operatorname*{argmin}_{\gamma \in X} E(\gamma),$$

where $E : X \to \mathbb{R}$ is a scalar loss function and γ^* is the trajectory that minimizes it. The optimization imperative is to find the best trajectory in the unrestricted space X: the trajectory with the lowest loss, wherever in X it may lie. This is the structure underlying gradient descent, free-energy minimization, reinforcement learning, and the vast majority of contemporary machine learning architectures.

Admissibility-based organization has a fundamentally different structure:

$$\gamma(t) \in \mathcal{A}(t) \quad \text{for all } t.$$

The admissibility imperative is not to minimize any scalar quantity but to remain within a structured region of trajectory space. There is no loss function being minimized; there is a constraint being maintained. The difference is not merely technical. Maintaining a constraint is categorically different from minimizing a loss because constraints are binary in a way that losses are not: a trajectory either satisfies the constraint or it does not, and violation of a viability constraint does not merely degrade performance but ends the organizational system entirely.

Proposition 6.5. *The admissibility imperative $\gamma(t) \in \mathcal{A}(t)$ is not expressible as the minimization of any continuous scalar loss function $E : X \to \mathbb{R}$ without fundamentally distorting the structure of the constraint.*

Proof. Suppose for contradiction that there exists a continuous $E : X \to \mathbb{R}$ such that $\operatorname{argmin}_\gamma E(\gamma) \subset \mathcal{A}(t)$ and such that E faithfully represents the admissibility constraint in the sense that trajectories near $\partial\mathcal{A}(t)$ have high E values while trajectories well within $\mathcal{A}(t)$ have low E values. For this to represent a genuine viability constraint, E must be discontinuous across $\partial\mathcal{A}(t)$: trajectories inside $\mathcal{A}(t)$ remain viable, while trajectories outside $\mathcal{A}(t)$ are organizationally extinct and have no finite loss value in any meaningful sense. But a continuous E cannot be discontinuous across $\partial\mathcal{A}(t)$. Therefore, any continuous loss function that approximates the admissibility constraint must either smooth out the constraint boundary — allowing trajectories outside $\mathcal{A}(t)$ to be treated as viable with high but finite cost — or treat the boundary as a soft penalty, which removes the categorical character of the constraint. Neither representation is faithful. The admissibility imperative is therefore not expressible as the minimization of any continuous scalar loss function without distorting its categorical structure. $\square$

This distinction has immediate consequences for the design of artificial systems aimed at genuine agency. Systems trained by loss minimization have optimization as their fundamental organizational principle: they are built to find the trajectory in X with the lowest loss, whether or not that trajectory respects any deeper organizational constraint. Systems organized around admissibility have persistence as their fundamental organizational principle: they are built to remain within $\mathcal{A}(t)$ rather than to minimize any scalar. An admissibility-based system can be worse at loss minimization than an optimization-based system while being more genuinely agentive, because its trajectories are constrained to remain within viable organizational space even when more accurate or higher-reward trajectories would exit it.

This opposition should not be read as eliminative. The point is not that optimization is useless or that prediction-error minimization is never appropriate, but that optimization is correctly understood as a local regulatory tool subordinated to a global admissibility imperative rather than as the constitutive principle of cognition. Within a given stratum $\mathcal{S}_i(t)$, optimization over a well-defined loss surface is a perfectly legitimate description of local dynamics. The error is to mistake this local description for the global ontology of the cognitive system — to assume that because local dynamics can be characterized as optimization, the global organizational imperative must therefore be optimization as well. The relationship is hierarchical: admissibility governs which strata are accessible, and within any given stratum, local optimization may accurately describe the microstructure of regulatory dynamics. A unified account must preserve both the local validity of optimization and the global priority of admissibility, rather than collapsing either into the other.

The distinction also clarifies a recurring tension in the alignment literature. Alignment by loss minimization — training a system to minimize the discrepancy between its outputs and human preferences — is an optimization operation. It locates the system's trajectory at the point in X with minimal preference-discrepancy loss, without any guarantee that this trajectory lies within a stable organizational region compatible with human viability over time. Alignment as shared admissibility maintenance is a different operation: it establishes a joint admissible trajectory space $\mathcal{A}_{\text{shared}}(t) \subset X_{\text{human}} \times X_{\text{system}}$ and requires that the system's trajectory remain within the shared admissible region. The two operations are not equivalent, and their divergence grows as the system's capabilities increase: a highly capable optimizer will find trajectories far outside the shared admissible space if loss minimization does not penalize such trajectories sufficiently, while a system genuinely organized around shared admissibility maintenance would refuse such trajectories categorically.

6.10 Admissibility and Enactivism: A Distinction

Readers familiar with enactivist and embodied cognitive science may be inclined to interpret the admissibility framework as a more formal presentation of positions already established in that tradition — as a mathematization of Varela, Noë, or Gibson's ecological approach. This reading, while understandable, misses what is structurally distinctive about the present framework.

Enactivism, as developed by Varela, Thompson, and Rosch [13], argues that cognition is constituted by the structural coupling between organism and environment and is not reducible to internal representational states. It emphasizes autopoiesis, sensorimotor contingency, and the co-constitution of agent and world. Gibson's ecological psychology [3] locates perception in the direct pickup of affordances from the optic array, bypassing inferential mediation. Both traditions correctly reject the representationalist picture and both correctly insist on the embodied, relational character of cognitive agency. The present essay has drawn on both.

The admissibility framework, however, is not merely an endorsement of these positions. It departs from standard enactivism in at least three respects. First, where enactivism tends to characterize structural coupling in terms of sensorimotor contingency and phenomenological experience, the admissibility framework centers a formal geometrical object — $\mathcal{A}(t) \subset X$ — and develops the consequences of its topology. The framework is therefore more tractable to formalization and more directly connected to mathematical tools such as dynamical systems theory, control-theoretic viability kernels, and persistence homology. Second, where enactivism grounds its account primarily in embodiment and sensorimotor engagement, the admissibility framework extends naturally to symbolic, institutional, and multi-scale temporal structures through the stabilization-operator account of the clipboard chapter. The same geometry that governs sensorimotor admissibility governs the admissibility of conceptual trajectories, institutional trajectories, and civilizational trajectories, because the framework is defined at the level of organizational topology rather than at the level of bodily experience. Third, the $\mathfrak{sl}_2$-style analysis of the preceding sections points toward a program of understanding global organizational coherence through the interaction of local generators: a program that is compatible with but not derivable from the enactivist tradition.

The question of how $\mathcal{A}(t)$ is empirically tractable is a legitimate challenge that the present essay does not fully resolve. Several directions are available. Viability kernels from control theory provide a formal analogue of admissible sets: the viability kernel of a constrained dynamical system is the largest set of initial conditions from which the system can remain within the constraint set indefinitely, and algorithms exist for computing it in tractable cases [15]. Metabolic phase boundaries and energetic viability surfaces define empirical admissibility boundaries in biological systems at the physiological level. Persistence homology, a tool from topological data analysis, provides methods for identifying stable topological features of organizational trajectories from time-series data without requiring explicit models of the underlying dynamical system. Ecological affordance graphs, following Gibson's tradition, provide a discrete approximation to the admissibility structure of an organism's perceptual environment. None of these is a complete implementation of the admissibility framework, but together they indicate that the framework is not merely metaphysical: it points toward an empirical research program whose formal tools are already partially available.

Chapter 7

Artificial Intelligence and the Limits of Prediction

7.1 Why LLMs Appear Intelligent

Large language models present a genuine puzzle for the theory of cognition. They exhibit behaviors that, under ordinary attributions, suggest understanding, reasoning, anticipation, and flexible response to novel situations. Yet they lack the embodiment, metabolism, ecological embedding, and organizational persistence that the present framework identifies as constitutive of genuine cognition. How can systems that are, on the analysis offered here, sophisticated compression engines appear to exhibit so many features of genuine intelligence?

The answer lies in a precise characterization of what LLMs actually do, and why that is impressive without being the thing it appears to be. An LLM is a system trained to approximate the conditional distribution of text continuations given text prefixes: given a sequence of tokens, it generates a probability distribution over possible next tokens. Through this training, it learns an extraordinarily rich compressed representation of the statistical regularities of human language use. The richness of this representation accounts for the system's apparent semantic competence: because human language use encodes vast amounts of structured information about the world, relationships, and reasoning patterns, a system that accurately models those patterns will exhibit behaviors that mimic the output of genuine understanding.

This is not a dismissal. The compression achieved by large-scale language modeling is a genuine scientific achievement, and the apparent semantic competence it produces is a genuine phenomenon. The point is that symbolic continuation is not equivalent to organismic agency, and the distinction matters for understanding what these systems can and cannot do.

LLMs are highly effective compression systems: they represent the statistical structure of enormous symbolic manifolds in a compact form that allows accurate continuation and interpolation. They are sophisticated symbolic continuation engines: they generate outputs that are locally coherent with their inputs and statistically consistent with the distributions they were trained on. They are powerful distributional stabilizers: they maintain coherence within the space of plausible continuations even when inputs are ambiguous, incomplete, or novel. These are remarkable capacities, and they produce outputs that are genuinely useful across an enormous range of tasks.

But the intelligence they exhibit is structurally different from the intelligence of an organism navigating an admissible trajectory space. The difference is not merely one of capability or scale; it is a difference in organizational type. An LLM does not maintain bodily continuity, negotiate metabolic constraints, preserve the conditions of its own self-repair, or manage the dynamical coupling with an ecological environment on which its continued functioning depends. It optimizes correspondence — the correspondence between its outputs and the statistical distribution of its training data — rather than persistence. And as the preceding chapters have argued, correspondence and persistence are not the same organizational imperative; they generate different kinds of systems, with different strengths, different limitations, and different failure modes.

7.2 Prediction Without Persistence

The most important limitation of current AI systems, from the perspective of the present framework, is not their failure to achieve human-level performance on specific benchmarks but their structural incapacity for organizational persistence. An LLM does not have an admissible trajectory space in the biological sense because it has no organizational integrity to preserve. Its outputs are not constrained by metabolic viability, ecological embedding, or the requirement to maintain the conditions of its own continued functioning. It can generate outputs that are locally coherent without the outputs being globally consistent with any organizational imperative beyond the statistical regularities of its training distribution.

This structural incapacity explains several features of current AI systems that are often treated as contingent engineering problems rather than principled limitations. The fragility of current systems under distribution shift — their tendency to fail badly when their inputs differ significantly from their training distribution — is not merely a consequence of insufficient training data but a consequence of the fact that the systems have no organizational coherence to fall back on when the familiar statistical patterns are absent. An organism navigating an unfamiliar environment is doing something fundamentally different from an LLM processing an out-of-distribution input: the organism is managing its interface with an unfamiliar environment while maintaining its organizational integrity; the LLM is extrapolating from a statistical model that may not apply.

The contrast can be stated as a pair of inequalities:

$$\text{Continuation} \neq \text{Agency}$$

and

$$\text{Statistical coherence} \neq \text{Organizational integrity}.$$

Current AI systems achieve continuation and statistical coherence to an impressive degree. They lack agency and organizational integrity in the senses defined in this essay. Whether this is a fundamental limitation or a contingent engineering problem is a question that cannot be answered on a priori grounds; it depends on whether the organizational properties that constitute genuine agency are achievable through the kind of statistical learning that current systems employ, or whether they require different organizational principles altogether.

The present framework suggests that the latter is the case. Genuine agency, as defined here, requires the maintenance of a stable admissible trajectory space under environmental perturbation, which in turn requires some form of organizational persistence: a substrate that has constraints to maintain, that can be disrupted by violations of those constraints, and that has resources for recovering from perturbation. Statistical learning systems that optimize correspondence rather than persistence lack these organizational properties by design. They can approximate the outputs of agentive systems, and this approximation can be extraordinarily accurate, but approximation is not implementation.

7.3 Toward Ecological Artificial Systems

What would an artificial system look like that instantiated genuine agency in the sense defined here? This is not primarily a question about computational architecture but about organizational principles. The framework developed in this essay suggests several requirements that any such system would need to satisfy.

First, constraint preservation. The system would need to have genuine organizational constraints — structural conditions on its operation whose violation disrupts its functioning — not merely optimization objectives whose violation degrades its performance. A constraint in the relevant sense is a condition that the system must maintain in order to continue operating, not merely a condition that it is rewarded for maintaining. The difference is the difference between a metabolic requirement and an objective function: the former admits of no trade-off, the latter is always balanced against other objectives.

Second, dynamical embodiment. The system would need to be embedded in a physical environment with which it maintains ongoing dynamical coupling: its internal dynamics would need to be shaped by environmental structure in ways that are not mediated by explicit inference over stored representations. This is not merely a requirement for physical instantiation but a requirement for the kind of anticipatory entrainment that characterizes biological cognition.

Third, admissibility navigation. The system would need to navigate a structured admissible trajectory space in ways that preserve organizational integrity rather than merely optimizing an objective function. This requires that the system's action-generation architecture be organized around the management of admissibility constraints rather than the maximization of a reward signal.

Fourth, multi-scale ecological coupling. The system would need to maintain organizational coherence across multiple temporal and spatial scales simultaneously: it would need to be coupled to its environment at the scale of immediate sensorimotor interaction, at the scale of longer-term planning and organization, and at the scale of the ecological niche within which it is embedded. These couplings would need to be mutually consistent and jointly maintained, rather than optimized independently.

These requirements are demanding, and it is not obvious that they can be satisfied by any architecture within the current AI paradigm. But they point toward a research agenda that goes beyond the scaling of prediction systems toward a more fundamental reconceptualization of what artificial intelligence is for. If the goal is to build systems that are genuinely agentive in the sense defined here, rather than systems that produce impressive approximations of agentive behavior, then prediction-error minimization is not sufficient, and the organizational principles of biological cognition provide a more appropriate target.

7.4 A Comparative Summary of the Two Frameworks

The argument developed across this essay can be consolidated by articulating the two frameworks along five dimensions of contrast. The purpose is not to replace the preceding analysis with a schema but to make the structural differences perspicuous in a form that facilitates precise disagreement.

The first dimension is the primary imperative. Prediction-error frameworks hold that the organism's fundamental organizational aim is correspondence: the maintenance of an internal generative model that accurately anticipates sensory input and minimizes the discrepancy between prediction and observation. The admissibility framework holds that the organism's fundamental organizational aim is persistence: the maintenance of the dynamical conditions necessary for continued existence as a viable organized system. Correspondence-seeking and persistence-seeking overlap in favorable environments, but they are not identical operations and they diverge under pressure. When correspondence-seeking would require trajectories that exit $\mathcal{A}(t)$, the organism prioritizes persistence; when persistence requires tolerating predictive inaccuracy, the organism tolerates it. The empirical prevalence of motivated reasoning, perceptual constancy effects, and the stabilization of functional states under sensory deprivation all reflect the priority of persistence over correspondence.

The second dimension is mathematical structure. Prediction-error frameworks are optimization frameworks: they characterize cognition as the minimization of a scalar loss function over a trajectory space, with free energy, prediction error, or reward discrepancy serving as the quantity to be minimized. The admissibility framework is a constraint-navigation framework: it characterizes cognition as the maintenance of a trajectory within a structured admissible region, with the constraint being categorical rather than scalar. As established in Proposition 6.5, the admissibility imperative cannot be faithfully expressed as a continuous loss minimization without distorting the categorical character of the viability boundary. The two mathematical structures generate different failure modes: optimization systems fail by finding minima outside the viable region, while admissibility systems fail by stratum descent below the boundary of their current organizational stratum.

The third dimension concerns the organism-environment relationship. Prediction-error frameworks characterize the organism's relationship to its environment primarily in terms of inference: the organism maintains an internal model of environmental structure and updates it through Bayesian inference on prediction errors. The admissibility framework characterizes the organism's relationship to its environment primarily in terms of entrainment and interface management: the organism's internal dynamics are physically coupled to environmental structure in ways that produce adaptive reorganization without requiring explicit inference over stored models, and cognition is the ongoing regulation of the organism-environment interface $\partial O(t)$ rather than the updating of an internal model.

The fourth dimension is the relationship between local and global structure. Prediction-error frameworks treat local inference as constitutive of global cognition: the organism's cognitive organization is identified with its local predictive machinery, and global behavior is the aggregate output of local inferential operations. The admissibility framework treats local operations as tangent approximations to a global organizational manifold: local predictive operators $P_t : T_xX \to T_{x+\delta}X$ are regulatory tools subordinated to the global admissibility structure $\mathcal{A}(t) \subset X$, and understanding the local operators does not determine the global topology any more than knowing the tangent bundle of a manifold determines its global topological type.

The fifth dimension is the account of organizational states and transitions. Prediction-error frameworks treat the organism's state space as a continuous loss surface over which gradient-based dynamics operate; different states differ quantitatively in their loss values. The admissibility framework treats the organism's state space as a stratified space with qualitatively distinct organizational strata separated by singular boundaries; different strata differ in the organizational modes they support and the attractors they make available. Pathological transitions are stratum descents that require boundary reconstruction for recovery, not local error corrections that move the organism toward a better point on a continuous loss surface.

These five contrasts together constitute what the reviewer of an earlier draft correctly characterized as a difference of ontological substrate rather than a difference of theoretical emphasis. The admissibility framework is not a refinement of prediction-error frameworks but a replacement of their underlying ontological commitments — a replacement that preserves their local validity while relocating them within a more adequate account of the global organizational imperative of living systems.

Conclusion

This essay has argued that cognition cannot be reduced to prediction-error minimization, that compression is not equivalent to organization, and that the dominant computational and predictive frameworks in neuroscience and artificial intelligence have systematically inflated the scope of their formalisms beyond what the formalisms can legitimately support. The reification error — the inference from the effectiveness of a mathematical description to the conclusion that the described system literally implements the description's structures — has produced a neuroscience that is substantially an autobiography of its own methodology, and an AI discourse that mistakes sophisticated statistical approximation for genuine intelligence.

The alternative framework developed here centers on three related reconceptualizations. First, the shift from prediction to admissibility: cognition is not primarily the minimization of discrepancy between predictions and observations but the navigation of a structured admissible trajectory space in ways that preserve organizational integrity. The organism's central problem is not epistemic — it is not primarily trying to construct an accurate world-model — but organizational: it is trying to remain within the region of trajectory-space in which continued existence as an organized system is possible.

Second, the shift from storage to stabilization: memory and symbolic cognition are not primarily storage-and-retrieval operations but organizational stabilization processes. The structures that cognitive systems produce — essays, notes, proofs, rituals, institutions, narratives — function as topological anchors that preserve the availability of organizational trajectories across temporal gaps and perturbations. They are not representations of content but constraints on possibility: they carve out and preserve regions of the admissible trajectory space.

Third, the shift from logic to topology: the fundamental structure of cognition is not truth-functional computation but coherence-preserving deformation within an admissible organizational space. Cognition is the management of the organism-environment interface in ways that maintain topological coherence across dynamic transformations. It is not primarily a propositional operation but a geometrical one.

These three shifts can be summarized in three pairs of inequalities, which collect the essay's central claims:

$$\text{Representation} \longrightarrow \text{Organization}$$

$$\text{Computation} \longrightarrow \text{Constraint Navigation}$$

$$\text{Prediction} \longrightarrow \text{Persistence}$$

The arrows here do not indicate logical entailment but conceptual progression: the terms on the right are not eliminations of the terms on the left but reconceptualizations that preserve what is genuine in the original terms while locating them within a more adequate organizational framework. Organisms predict, represent, and compute. But they do so as subordinate operations within a more fundamental organizational imperative: the maintenance of coherent admissible trajectories within dynamically evolving ecological fields.

This conclusion is not a refutation of computational neuroscience or artificial intelligence but a challenge to their self-understanding. The formalisms of prediction error, free energy minimization, and information processing are genuine tools with genuine scope. The challenge is to understand that scope accurately: to recognize what these formalisms capture and what they systematically miss, and to develop theoretical frameworks adequate to the full complexity of the phenomena they aspire to explain.

The line that perhaps best captures what remains missing from the dominant frameworks comes from Brette's argument: the real challenge is to explain constancy in a world where nothing remains static. This is the challenge that the admissibility framework attempts to meet. Not the challenge of accurate prediction, not the challenge of optimal compression, not the challenge of minimal surprise — but the challenge of remaining the same kind of thing, in the same kind of world, under the continuous pressure of an environment that is never perfectly accommodating and never entirely alien. That is the problem of living intelligence, and it is not yet the problem that our most sophisticated artificial systems have been built to solve.

Appendix A

Variational Geometry of Admissible Trajectories

The admissibility framework developed in the main text can be formalized through a constrained variational geometry in which viable cognitive organization is represented not by optimization over arbitrary trajectories but by persistence-preserving flow within a restricted manifold.

Let X denote the full ecological trajectory manifold of the organism and let $\mathcal{A}(t) \subset X$ denote the admissible region at time t. A trajectory $\gamma : [0, T] \to X$ is admissible if and only if $\gamma(t) \in \mathcal{A}(t)$ for all $t \in [0, T]$.

Rather than defining cognition through minimization of a scalar loss functional, we define organizational persistence through the constrained action

$$\mathcal{S}[\gamma] = \int_0^T L(\gamma(t), \dot{\gamma}(t), t)\, dt$$

subject to the viability condition $\Phi(\gamma(t), t) \geq 0$, where

$$\Phi : X \times \mathbb{R} \to \mathbb{R}$$

defines the admissibility boundary. The admissible trajectory problem therefore becomes $\delta\mathcal{S} = 0$ subject to $\gamma(t) \in \mathcal{A}(t)$.

Unlike unconstrained optimization systems, admissible systems exhibit singular boundary behavior. The Euler–Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\gamma}}\right) - \frac{\partial L}{\partial \gamma} = 0$$

hold only within the interior $\mathrm{Int}(\mathcal{A}(t))$. At the admissibility boundary $\partial\mathcal{A}(t)$, additional geometric conditions emerge:

$$\langle \dot{\gamma},\, \nabla\Phi \rangle \geq 0.$$

This inequality states that viable trajectories cannot possess velocity vectors directed outward through the admissibility boundary. Biological cognition is therefore not free optimization but constrained geodesic navigation within a time-dependent viability geometry.

One may define the admissibility metric tensor $g_{ij}(x,t)$ such that distances correspond not to Euclidean separation but to transition difficulty between organizational states. The induced line element becomes

$$ds^2 = g_{ij}(x,t)\, dx^i dx^j.$$

Near pathological regions, the metric becomes highly anisotropic,

$$\lambda_{\max}(g) \gg \lambda_{\min}(g),$$

producing elongated attractor valleys and constrained mobility through trajectory space. Depression, addiction, and rigidity can therefore be modeled as curvature singularities in the admissibility metric itself.

The Ricci scalar $R(x,t)$ then acquires organizational interpretation: positive curvature corresponds to compressive attractor concentration, while negative curvature corresponds to expansive exploratory freedom. Accordingly, $R(x,t) \to +\infty$ corresponds to pathological trajectory collapse, whereas $R(x,t) \ll 0$ corresponds to excessive organizational diffusion and loss of coherence. The organism's task is therefore neither maximal compression nor maximal exploration but bounded curvature maintenance within a navigable admissibility geometry.

Appendix B

Categorical Structure of Organizational Persistence

The admissibility framework admits a natural categorical interpretation in which organizational systems are represented as objects preserving coherence under constraint-preserving morphisms.

Let **Adm** be the category whose objects are admissibility structures $\mathcal{A}_i$ and whose morphisms $f : \mathcal{A}_i \to \mathcal{A}_j$ preserve organizational viability: $f(\gamma) \in \mathcal{A}_j$ for all $\gamma \in \mathcal{A}_i$. A morphism is therefore not merely a transformation but a persistence-preserving organizational transport operation. Composition satisfies $(g \circ f)(\gamma) = g(f(\gamma))$, and preservation of admissibility implies $g \circ f : \mathcal{A}_i \to \mathcal{A}_k$.

The central distinction between predictive systems and persistence systems becomes categorical. Prediction systems optimize over state descriptions, $P : X \to M$. Persistence systems preserve admissible structure, $F : \mathbf{Adm} \to \mathbf{Adm}$. Cognition is therefore not fundamentally representational but functorial.

A stabilization operator of the kind discussed in Chapter 5 becomes a functor $S : \mathbf{Adm} \to \mathbf{Adm}$ preserving coherence across temporal discontinuities. An essay, proof, or institution acts as a coherence-preserving transport mechanism between admissible regions: $S(U_i) \cong U_j$. Identity itself becomes a recursively stabilized endofunctor $I : \mathbf{Adm} \to \mathbf{Adm}$ satisfying $I \circ I \simeq I$. Selfhood therefore corresponds not to substance persistence but to recursive constraint-preserving self-mapping.

The pathological breakdown of selfhood can be expressed as functorial failure: $I_{t+1} \not\simeq I_t$, meaning the system can no longer transport its organizational coherence through time. This categorical formulation clarifies why symbolic systems scale cognition. Language, mathematics, institutions, and rituals preserve not information but morphism stability: they preserve the existence of admissible transformations between organizational regions. Civilization itself may be interpreted as a distributed higher-order sheaf of stabilization functors preserving coherence across temporally separated agents.

Appendix C

Spectral Decomposition of Cognitive Organization

The topology of admissibility can be analyzed spectrally through operators acting over organizational manifolds.

Let $\mathcal{H} = L^2(X)$ be the Hilbert space of organizational states over the trajectory manifold X. Define the admissibility operator $\hat{\mathcal{A}}$ acting on state functions $\psi(x,t)$ such that $\hat{\mathcal{A}}\psi = \lambda\psi$. Eigenfunctions correspond to dynamically stable organizational modes, while eigenvalues correspond to persistence stability. Large positive eigenvalues correspond to highly coherent attractor structures $(\lambda_i \gg 0)$, while negative eigenvalues correspond to unstable or decaying organizational modes $(\lambda_i < 0)$.

The full organizational state decomposes spectrally:

$$\psi(x,t) = \sum_i c_i(t)\, \psi_i(x).$$

Cognition is therefore not a point-state process but a dynamically evolving spectral superposition of organizational modes. The temporal coefficients $c_i(t)$ represent varying activation strengths of persistence structures across time. Healthy cognition corresponds to bounded spectral coherence,

$$\sum_i |c_i(t)|^2 < \infty,$$

together with stability of dominant organizational modes. Pathology emerges when spectral concentration collapses excessively, $|c_k(t)| \to 1$ for a narrow pathological mode, or when coherence fragments chaotically across incompatible modes, $\sum_i |c_i(t)|^2 \to \infty$.

This interpretation connects naturally to dynamical systems theory. The Laplace–Beltrami operator Δ_g over the admissibility manifold generates diffusion dynamics:

$$\frac{\partial \psi}{\partial t} = \Delta_g \psi - V(x)\psi,$$

where $V(x)$ acts as an organizational potential landscape. Stable cognitive identities correspond to metastable spectral basins: localized eigenmodes resistant to perturbation.

The spectral interpretation also clarifies the LLM situation developed in Chapter 7. Large language models approximate projections of the dominant symbolic eigenmodes of human cultural manifolds:

$$\psi_{\text{LLM}} \approx \Pi(\psi_{\text{human}}).$$

But projection onto dominant symbolic modes is not equivalent to maintaining the admissibility geometry generating those modes originally. The distinction between symbolic continuation and genuine agency is therefore the distinction between approximating eigenfunctions of a persistence manifold and maintaining the manifold itself:

$$\text{spectral approximation} \neq \text{organizational persistence}.$$

Bibliography

[1] Romain Brette, *The Brain, In Theory*, Princeton University Press, 2026.

[2] Romain Brette, *Predictive Coding is Not a Theory of Anticipation*, 2026.

[3] James J. Gibson, *The Ecological Approach to Visual Perception*, Routledge, 1979.

[4] Henri Bergson, *Matter and Memory*, Zone Books, 1990.

[5] Alfred North Whitehead, *Science and the Modern World*, Macmillan, 1925.

[6] Andy Clark, *Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science*, *Behavioral and Brain Sciences* 36 (2013), 181–204.

[7] Karl Friston, *The Free-Energy Principle: A Unified Brain Theory?*, *Nature Reviews Neuroscience* 11 (2010), 127–138.

[8] Rajesh P. N. Rao and Dana H. Ballard, *Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-Classical Receptive-Field Effects*, *Nature Neuroscience* 2 (1999), 79–87.

[9] Jerry A. Fodor, *The Language of Thought*, Harvard University Press, 1975.

[10] Allen Newell and Herbert A. Simon, *Human Problem Solving*, Prentice-Hall, 1972.

[11] Norbert Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine*, MIT Press, 1948.

[12] Gilbert Simondon, *Individuation in Light of Notions of Form and Information*, University of Minnesota Press, 2020.

[13] Francisco J. Varela, Evan Thompson, and Eleanor Rosch, *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press, 1991.

[14] James E. Humphreys, *Introduction to Lie Algebras and Representation Theory*, Springer, 1972.

[15] Jean-Pierre Aubin, *Viability Theory*, Birkhäuser, 1991.