EPISTEMIC REDUCTION AND ADMISSIBLE PROJECTION

Anderson's Epistemology as a Theorem of RSVP Field Theory

Flyxion
Independent Researcher

Standard Galactic Edition

Monica Anderson's original texts are cited and quoted with full attribution. The formal developments, theorems, and RSVP machinery are original contributions of the present author.

Typeset with LuaLaTeX using TeX Gyre Pagella.

CONTENTS

PREFACE

You cannot Reason about that which you do not Understand.

Monica Anderson

This monograph has a precise thesis: Monica Anderson's epistemological program is derivable from the geometry of admissible projection and RSVP field theory, not merely compatible with it.

The derivation proceeds through the admissibility quotient construction. Once a trajectory space $\mathcal{X}$ and a continuous admissibility operator $\mathcal{A}$ are in hand, Anderson's distinction between Understanding and Reasoning emerges automatically as the difference between constructing and operating within a representational manifold. Her Priority Thesis is a structural tautology about the dependency between projections and operators-on-projections. Her Salience criterion becomes admissibility sensitivity. Her Corpus Congruence becomes the level-set structure of a smoothed familiarity potential field. Her critique of classical AI becomes the Projection Supremacy Failure Theorem. Her Wisdom Salon becomes a sheaf-theoretic colimit whose global section is guaranteed by the descent axiom. Calvin's uphill river becomes a monotone chain of admissibility quotients with decreasing mutual information.

None of these require Anderson's framework as an assumption.

The monograph is mathematically candid about the status of each result. Six categories of claim appear. Theorems are proved from the stated hypotheses. Propositions are theorems with lighter hypotheses or shorter proofs. Lemmas are auxiliary results used in other proofs. Corollaries follow immediately from preceding results. Conjectures are plausible claims for which a proof route is described but not completed; they are explicitly labelled and the missing ingredients are identified. Remarks interpret and contextualize.

The principal remaining conjectural element is the bounded curvature characterization of admissibility-compatible learning: the Local Extension Stability Theorem provides a genuine one-directional result, but the biconditional that the previous draft asserted has been downgraded to a conjecture pending further geometric development. Readers who find the conjecture interesting will find the missing ingredients identified clearly in Chapter **??**.

Anderson is named in the subtitle as an intellectual predecessor. This monograph provides the proofs she did not supply, with the honesty their subject requires.

Part I

Anderson’s Epistemological System

CHAPTER 1

INTELLIGENCE, UNDERSTANDING, AND REASONING

Intelligence = Understanding + Reasoning.

Monica Anderson, Why AI Works

1.1 The Central Decomposition

Anderson's framework opens with a functional decomposition of intelligence into two irreducible faculties. Understanding is subconscious, intuitive, model-free, holistic, expensive, and fallible in a characteristic way: its errors are graceful near-misses rather than catastrophic failures at arbitrary inputs. Reasoning is conscious, logical, model-based, cheap per inference step, and conditionally infallible given correct input within a specified model space.

The central structural claim, stated explicitly by Anderson and repeated in multiple forms, is that Understanding is *prior to* Reasoning. One cannot reason about a domain one has not yet understood. The converse does not hold. This asymmetry Anderson calls the Priority Thesis. Part III derives it as Theorem **??**.

1.2 Properties of the Two Faculties

Understanding is subconscious (no introspective access to its operation), intuitive (rapid and parallel, answering without deliberation), model-free (no a priori symbolic representation of the problem domain), holistic (context is taken as a whole), expensive (the majority of a brain's processing budget), and fallible. Reasoning is conscious, sequential, model-based, cheap per inference step, and conditionally infallible. The guarantee is real but narrow: it applies only inside the model space from which context has been deliberately stripped.

Anderson draws on Kahneman's System 1 / System 2 distinction but diverges in one important respect. Kahneman treats the two systems as interacting processes within a single cognitive architecture. Anderson treats them as requiring fundamentally different computational paradigms with different failure modes.

1.3 The Priority Thesis

Proposition 1.1 (Anderson's Priority Thesis, informal)**.** *Understanding is strictly prior to Reasoning. Understanding does not presuppose Reasoning. Reasoning presupposes Understanding.*

This generates three consequences for AI methodology. First, any architecture that attempts to produce understanding through symbolic reasoning inverts the dependency. Second, twentieth-century AI (build a model, reason over it) attacked the second step while bypassing the first. Third, the correct question about AI architecture is epistemological.

Part III proves Proposition **??** as Theorem **??**.

CHAPTER 2

MODELS, REDUCTIONISM, AND THE PATHOLOGY OF PROJECTION

Model Based Problem Solving is the greatest invention our species has ever made.

Monica Anderson, Our Greatest Invention

2.1 What a Model Is

Definition 2.1 (Model, after Anderson)**.** A Model is a simplification of reality produced by discarding context in order to make a fragment of the world tractable to formal manipulation. Models include scientific theories, equations, statistical models, computer programs, and naive personal simplifications.

Definition 2.2 (Reductionism, after Anderson)**.** Reductionism is the use of Models.

Models are the most productive intellectual tools humanity has produced. A Model stripped of all domain-specific detail can be applied by anyone trained in it, independently of whether they have any direct experience with the original problem context. Newton's $F = MA$ works in every context in which mass, force, and acceleration are well-defined precisely because every other detail has been discarded. The scientific enterprise of the past four centuries is largely a library of such reusable compressions.

To create a Model is to perform a compression. The scientist surveys a domain, identifies the features that determine the phenomenon of interest, discards the rest, and publishes the result as a context-free representation that others can apply without repeating the original investigation. Anderson's central observation is that this compression is not a description of a process internal to science. It is a pre-scientific operation that science presupposes. The decision about which features to retain and which to discard must be made by an agent who already understands the domain at a level that precedes the Model. The Model is the output of Understanding, not its vehicle.

2.2 The Model–Projection Correspondence

The relationship between Anderson's Models and the formal machinery of Part II is exact rather than metaphorical.

Proposition 2.3 (Models are Projections)**.** *In the framework of Part II, a Model in Anderson's sense corresponds to a projection $\pi : \mathcal{X} \to \mathcal{M}$ from a trajectory space $\mathcal{X}$ to a representational manifold $\mathcal{M}$. The act of Model creation is the act of specifying π. The context-free content of the Model is the structure of $\mathcal{M}$. The discarded context is the information in the fibers $\pi^{-1}(m)$ for each $m \in \mathcal{M}$.*

This correspondence makes Anderson's claims about Models testable as claims about projections. Model reusability becomes the claim that $\mathcal{M}$ supports correct reasoning independently of which $x \in \pi^{-1}(m)$ was the original input. Model applicability becomes the claim that $\pi^{-1}(\mathcal{M})$ covers the region of $\mathcal{X}$ where the Model is being used. Model failure becomes the claim that the input x lies in $\mathcal{X} \setminus \pi^{-1}(\mathcal{M})$.

2.3 Compression, Utility, and the Utility Threshold

The compression effected by a Model is not a loss but a gain, under the right conditions. A Model that compresses a 10^6-dimensional state space to a three-parameter equation does not impoverish the analyst; it enables the analyst to solve problems that would be computationally impossible without the compression. The utility of a Model is precisely a function of its compression ratio together with the accuracy of the admissibility structure it preserves.

Definition 2.4 (Model Utility)**.** The utility of a projection $\pi : \mathcal{X} \to \mathcal{M}$ for a class of tasks $\mathcal{T}$ is the product of its compression (the reduction in representational complexity $C(\mathcal{X}) - C(\mathcal{M})$) and its admissibility fidelity (the degree to which $\mathcal{A}_{\mathcal{M}}(\pi(x))$ approximates $\pi(\mathcal{A}(x))$ for x drawn from the task distribution). A Model is useful when high compression and high admissibility fidelity can be achieved simultaneously.

The scientific enterprise can be understood as the sustained search for projections with high utility: cases where the world happens to admit compressions that preserve enough admissibility structure to support reliable inference and action. Physics is the discipline where such compressions are most available. The life sciences, social sciences, and natural language are domains where they are most difficult to find, which is why those fields resist the kind of mathematical formalization that has been so productive in physics.

2.4 The Pathology: Projection Supremacy

The pathology Anderson calls Reductionism is not the use of Models. It is a specific failure mode that occurs when Models are mistaken for the reality they compress. Anderson's target is not compression as such but the cognitive error of treating the projection as complete: acting as though $\mathcal{M} = \mathcal{X}$.

This error has a precise name in the framework of Part II: Projection Supremacy (Definition **??**). A system in a state of Projection Supremacy applies its reasoning operator $R : \mathcal{M} \to \mathcal{M}$ to inputs x with non-zero projection error $d_H(\mathcal{A}(x), \mathcal{A}_{\mathcal{M}}(\pi(x))) > 0$, without any mechanism for detecting or compensating for this error.

The error is not that the system has a Model. The error is that the system has no representation of the gap between the Model and the world.

Proposition 2.5 (The Model–Projection Supremacy Chain)**.** *The following sequence describes the generation and eventual failure of a Reductionist cognitive system.*

1. ***Model creation.*** *A scientist performs Epistemic Reduction: $\pi : \mathcal{X} \to \mathcal{M}$ is constructed, discarding non-salient features. The result is a useful compression with well-defined applicability conditions.*

2. ***Institutionalization.*** *The Model is published, taught, and applied. Users learn $R : \mathcal{M} \to \mathcal{M}$ (the reasoning procedures of the field) without learning the construction of π. The original Understanding that produced the Model is separated from the Model itself.*

3. ***Scope inflation.*** *The Model is applied in contexts increasingly distant from its original derivation. The applicability conditions are either unknown to the users or disregarded. The projection error $d_H(\mathcal{A}(x), \mathcal{A}_{\mathcal{M}}(\pi(x)))$ grows.*

4. ***Projection Supremacy.*** *The system has no representation of the projection error. It applies R uniformly, producing confident outputs even in regions where $\pi^{-1}(\mathcal{M})$ does not cover x.*

5. ***Brittleness.*** *Catastrophic failure occurs at the edges of competence by Theorem* **??**. *The failure is unmodulated because the system has no internal signal for the growing discrepancy between $\mathcal{M}$ and $\mathcal{X}$.*

Remark 2.6. Proposition **??** is not a theorem; it is a phenomenological description of a pattern that Proposition **??** and Theorem **??** together explain. The sequence makes the transition from Anderson's critique of classical AI to the RSVP formal treatment legible: Anderson is describing the middle three steps of the chain (institutionalization, scope inflation, Projection Supremacy) in operational terms. The RSVP framework provides the geometry that explains why the fifth step is structurally inevitable once the fourth is in place.

2.5 Why Models Cannot Produce Understanding

Anderson's argument that Models cannot produce Understanding is the circularity corollary stated below. But in the expanded framework above, a more precise version can be given.

Proposition 2.7 (Irreducibility Barrier). *There exists a class of problem domains, which Anderson calls Bizarre Domains, in which no projection $\pi : \mathcal{X} \to \mathcal{M}$ can be simultaneously tractable* $(\dim \mathcal{M} \ll \dim \mathcal{X})$ *and admissibility-faithful* (*projection error below any fixed threshold*), *because the admissible future structure of $\mathcal{X}$ is not well-approximated by any low-dimensional quotient. Reductionist methods are structurally blocked in these domains.*

Proof sketch. A domain is Bizarre in Anderson's sense when the phenomena of interest are emergent: they depend on the entire configuration of the system and not on any low-dimensional projection of it. Formally, this means the equivalence classes of $\sim_{\mathcal{A}}$ are singletons or near-singletons: almost every state has a unique admissible future set. In this case the admissibility quotient $\mathcal{M}_0 = \mathcal{X}/{\sim_{\mathcal{A}}}$ is essentially isomorphic to $\mathcal{X}$ itself, and no compression of substantial ratio is admissibility-faithful. ☐

Corollary 2.8 (Model-Based AI Circularity). *Any AI architecture that attempts to produce Understanding through hand-crafted Models is self-defeating. Understanding is the construction of $\pi : \mathcal{X} \to \mathcal{M}$. A hand-crafted Model is a fixed π, specified in advance of the data. By Theorem* **??**, *any feature discarded by this fixed π cannot become salient through reasoning operations internal to $\mathcal{M}$. The architecture can therefore never discover admissibility-relevant features that its designers did not already know mattered.*

Remark 2.9. The corollary is stronger than a circularity argument. It is an impossibility result: a hand-crafted Model not only fails to generate Understanding for the features it retains; it structurally prevents the discovery of salience for the features it discards. This is the information-theoretic content of Anderson's claim that "models cannot create understanding": not only is there a circular dependency, but the projection forecloses the discovery of its own inadequacy.

CHAPTER 3

HOLISM, EPISTEMIC REDUCTION, AND BIZARRE DOMAINS

> Holism is the avoidance of Models.
>
> *Monica Anderson,* Two Dirty Words

Definition 3.1 (Holism, after Anderson)**.** Holism is the meta-strategic avoidance of a priori Models of the problem domain. A Holistic system operates directly in the problem domain without first constructing a symbolic representation.

Definition 3.2 (Epistemic Reduction, after Anderson)**.** Epistemic Reduction is the operation that discovers higher-level abstractions in lower-level data by discarding everything recognized as irrelevant at the current level of understanding.

Definition 3.3 (Salience, after Anderson)**.** A feature of an input state is salient if its retention is necessary for the current cognitive task. Salience is task-relative, context-sensitive, and level-specific.

Salience cannot be evaluated at the level of the feature itself: evaluating salience requires a higher-level context that already knows what the task demands. This is why deep architectures are necessary — the salience of a low-level feature can only be determined from a higher-level perspective the low-level layer does not yet possess.

3.1 Bizarre Domains

Anderson's 2007 *Artificial Intuition* introduced a taxonomy of problem domains that resist Logic-based analysis. She called these *Bizarre Domains*, adopting a term from Stephen Kercel, and identified four categories of obstructing properties: Chaos, Holism, Ambiguity, and Emergence. A Bizarre Domain, in Anderson's definition, contains at least one instance of each.

The taxonomy has often been read as a list of difficulties. In the admissibility geometry of Parts II and III, it is something more precise: a classification of the distinct ways in which projections fail.

Proposition 3.4 (Bizarre Domains as Projection Failure Modes)**.** *Anderson's four categories of Bizarre Domain properties correspond to four distinct failure modes of admissibility-preserving projection.*

1. ***Chaos*** *(sensitive dependence on initial conditions): the admissible future cone $\mathcal{A}(x)$ varies discontinuously with x, so nearby states may have radically different admissible futures. A projection π that is continuous on $\mathcal{X}$ cannot preserve discontinuous admissibility structure; projection error $d_H(\mathcal{A}(x), \mathcal{A}_{\mathcal{M}}(\pi(x)))$ grows rapidly outside the training distribution. Formally: the admissibility equivalence relation $\sim_{\mathcal{A}}$ has dense singleton classes, and the quotient $\mathcal{M}_0 \approx \mathcal{X}$ is uncompressible.*

2. ***Holism*** *(irreducibility, context-dependence): the admissible future cone $\mathcal{A}(x)$ depends on the full state $x \in \mathcal{X}$ rather than on any low-dimensional projection of it. Any projection $\pi : \mathcal{X} \to \mathcal{M}$ with $\dim \mathcal{M} \ll \dim \mathcal{X}$ necessarily violates admissibility preservation: the fiber $\pi^{-1}(\pi(x))$ contains states with distinct admissible futures. By Proposition* **??***, the domain is Bizarre in Anderson's holistic sense precisely when $\mathcal{M}_0 \approx \mathcal{X}$.*

3. ***Ambiguity*** *(multiple valid interpretations): multiple admissible continuations coexist from a single state. The admissible future cone $\mathcal{A}(x)$ is not a singleton: there are genuine forks in the accessible trajectory space. A Logic-based system, designed around unique correct answers, is structurally incapable of representing $|\mathcal{A}(x)| > 1$; it must prematurely collapse the ambiguity. Anderson's prescription (maintain multiple interpretations in parallel) corresponds to preserving the full measure $\mu(\mathcal{A}(x))$ rather than selecting a single element.*

4. ***Emergence*** *(system-level properties not recoverable from components): the admissibility operator $\mathcal{A}$ is not decomposable into component-wise admissibility operators. If $x = (x_1, \ldots, x_n)$, then $\mathcal{A}(x) \neq \mathcal{A}_1(x_1) \times \cdots \times \mathcal{A}_n(x_n)$; the accessible futures of the system cannot be computed from the accessible futures of its parts. This is the formal version of the claim that emergent properties "disappear when the system is taken apart."*

Remark 3.5. The four categories are not independent in Anderson's taxonomy: she notes they tend to co-occur in the domains where intelligence is most needed. In the admissibility framework this is also expected: a domain that is genuinely holistic (failure mode 2) will also tend to exhibit sensitive dependence (failure mode 1) because the global context that determines admissibility can be disrupted by small local changes. The failure modes interact because they all have a common source: the non-decomposability of the admissibility operator.

Anderson's diagnosis — that Logic-based AI fails in Bizarre Domains while Holistic methods succeed — is now a corollary of the geometry. Logic-based AI imposes a fixed projection before encountering the data. In Bizarre Domains, no fixed projection can preserve admissibility (failure modes 1 and 2) or represent the full structure of accessible futures (failure modes 3 and 4). Holistic methods defer the projection and learn it from the data, approximating the admissibility quotient rather than imposing a hand-authored one.

3.2 Salience and the Uphill River

Anderson adopts William Calvin's image of a river flowing uphill to characterize the directionality of Reduction. Physical causation flows downhill: from simpler descriptions to their consequences. Reduction flows uphill: from low-level data to high-level categories that did not exist before the reduction discovered them. In Part III, the uphill river is identified with the admissibility quotient construction, and formalized as a monotone chain of quotient spaces in the Recursive Abstraction Theorem (Theorem **??**).

The Bizarre Domain taxonomy now gives the uphill river a precise thermodynamic interpretation. A domain is Bizarre when the accessible future cone is not compressible. Moving uphill through the quotient hierarchy is the only strategy that can reduce representational complexity while preserving what matters, because any premature compression violates admissibility in at least one of the four modes. The river flows uphill precisely because the structure of accessible futures only becomes compressible at higher levels of abstraction.

CHAPTER 4

CORPUS CONGRUENCE, MACHINE UNDERSTANDING, AND THE AGI MYTH

Definition 4.1 (Corpus Congruence, after Anderson)**.** A Holistic system Understands an input to the degree that the input is congruent with the system's training corpus.

Anderson's core epistemological constraints:

1. *You can only learn what you almost already know.* (Patrick Winston.) New inputs must be adjacent to existing representations.
2. *All intelligences are fallible.*
3. *To detect novelty you must recognize everything familiar.*
4. *You cannot Reason about what you do not Understand.*
5. *You are known by the company you keep.* (Anderson's gloss on the Yoneda Lemma.)

Definition 4.2 (Artificial General Learner)**.** An Artificial General Learner is a system capable of learning to perform competently across wide problem domains through data exposure, without requiring antecedent domain-specific Understanding encoded as programs.

Part II

Geometric Machinery

CHAPTER 5

TRAJECTORY SPACES AND THE ADMISSIBILITY OPERATOR

5.1 Two Settings

The monograph operates in two settings, which must be kept distinct.

Metric setting. $(\mathcal{X}, d_{\mathcal{X}})$ is a compact metric space. The admissibility operator assigns compact subsets of $\mathcal{X}$, equipped with the Hausdorff metric. Salience, the quotient construction, and universality results live here. No Riemannian structure is assumed.

Smooth RSVP setting. $\mathcal{X}$ additionally carries a Riemannian metric g and a reference measure μ, and $\mathcal{A}(x)$ is a measurable subset varying smoothly with x. The accessibility entropy field, its gradient, Hessian computations, and the familiarity potential live here. Statements using ∇, ∇^2, or differential geometry apply only in this setting.

Definition 5.1 (Trajectory Space)**.** A trajectory space $(\mathcal{X}, d_{\mathcal{X}})$ is a compact metric space whose points $x \in \mathcal{X}$ are complete state histories of an agent embedded in its environment up to the current moment.

Definition 5.2 (Admissibility Operator)**.** An admissibility operator is a map $\mathcal{A} : \mathcal{X} \to 2^{\mathcal{X}}$ assigning to each state x a compact subset $\mathcal{A}(x) \subseteq \mathcal{X}$ of states accessible under the dynamical constraints of the agent-environment system. The operator is *Hausdorff-continuous* if $x \mapsto \mathcal{A}(x)$ is continuous in the Hausdorff metric d_H.

Definition 5.3 (Accessibility Entropy (smooth RSVP setting))**.** Let μ be a reference measure on $\mathcal{X}$. The accessibility entropy field is

$$S : \mathcal{X} \to \mathbb{R}, \qquad S(x) = \log \mu\big(\mathcal{A}(x)\big),$$

defined wherever $\mu(\mathcal{A}(x)) > 0$.

Definition 5.4 (Admissibility Equivalence)**.** $x \sim_{\mathcal{A}} y$ iff $\mathcal{A}(x) = \mathcal{A}(y)$ as subsets of $\mathcal{X}$.

Lemma 5.5 (Closure of the Equivalence Relation)**.** *In the metric setting, if $\mathcal{A}$ is Hausdorff-continuous then $\sim_{\mathcal{A}}$ is a closed equivalence relation on $\mathcal{X}$.*

Proof. Let $(x_n, y_n) \to (x, y)$ with $\mathcal{A}(x_n) = \mathcal{A}(y_n)$ for all n. Hausdorff-continuity gives $\mathcal{A}(x) = \lim_n \mathcal{A}(x_n) = \lim_n \mathcal{A}(y_n) = \mathcal{A}(y)$, so $x \sim_{\mathcal{A}} y$. □

Chapter 6

The Admissibility Quotient and Representational Manifolds

6.1 Two Notions of Admissibility-Preserving Projection

The admissibility operator $\mathcal{A}(x)$ returns a subset of $\mathcal{X}$, while the induced operator on a quotient or codomain returns a subset of that space. These cannot be compared as sets; an induced admissibility must be defined explicitly before any preservation condition can be stated.

Definition 6.1 (Induced Admissibility)**.** Given a surjection $\pi : \mathcal{X} \to \mathcal{M}$ and the admissibility operator $\mathcal{A}$ on $\mathcal{X}$, define the induced admissibility operator on $\mathcal{M}$ by

$$\mathcal{A}_{\mathcal{M}}(\pi(x)) \;:=\; \pi\big(\mathcal{A}(x)\big) \;\subseteq\; \mathcal{M}.$$

This is well-defined whenever $\pi(x) = \pi(y)$ implies $\pi(\mathcal{A}(x)) = \pi(\mathcal{A}(y))$.

Definition 6.2 (Admissibility-Preserving and Exact Minimal Projections)**.** Let $\pi : \mathcal{X} \to \mathcal{M}$ be a continuous surjection with induced admissibility as in Definition **??**.

1. π is *admissibility-preserving* if for all $x \in \mathcal{X}$:

   $$\pi\big(\mathcal{A}(x)\big) = \mathcal{A}_{\mathcal{M}}\big(\pi(x)\big).$$

   This asserts commutation with admissible dynamics. The projection may still separate states with identical admissible futures (retaining irrelevant information).

2. π is *exact minimal admissibility-preserving* if additionally:

   $$\pi(x) = \pi(y) \quad\Longleftrightarrow\quad \mathcal{A}(x) = \mathcal{A}(y).$$

   This collapses states if and only if they share identical admissible future sets.

Remark 6.3. The distinction is essential for the universal property theorem and the information minimality theorem. An admissibility-preserving projection in sense (1) may carry arbitrary additional information about x beyond its admissibility class. The theorems that follow state their hypotheses explicitly to avoid conflating these two notions.

6.2 The Quotient Construction

Theorem 6.4 (Existence of the Admissibility Quotient). *In the metric setting, suppose $(\mathcal{X}, d_{\mathcal{X}})$ is compact Hausdorff and $\mathcal{A}$ is Hausdorff-continuous. Then:*

1. *The quotient space $\mathcal{M}_0 = \mathcal{X}/{\sim_{\mathcal{A}}}$ is compact Hausdorff.*
2. *There exists a unique continuous surjection $\pi_{\mathcal{A}} : \mathcal{X} \to \mathcal{M}_0$ satisfying*

$$\pi_{\mathcal{A}}(x) = \pi_{\mathcal{A}}(y) \iff \mathcal{A}(x) = \mathcal{A}(y).$$

3. *$\pi_{\mathcal{A}}$ is exact minimal admissibility-preserving, with induced admissibility $\mathcal{A}_{\mathcal{M}_0}([x]) := \pi_{\mathcal{A}}(\mathcal{A}(x))$.*

Proof. By Lemma **??**, $\sim_{\mathcal{A}}$ is a closed equivalence relation. A closed equivalence relation on a compact Hausdorff space yields a compact Hausdorff quotient (standard quotient topology). The canonical projection $\pi_{\mathcal{A}} : x \mapsto [x]$ is continuous, surjective, and satisfies $\pi_{\mathcal{A}}(x) = \pi_{\mathcal{A}}(y)$ iff $[x] = [y]$ iff $\mathcal{A}(x) = \mathcal{A}(y)$ by definition of $\sim_{\mathcal{A}}$. The induced admissibility is well-defined on equivalence classes since $[x] = [y]$ implies $\mathcal{A}(x) = \mathcal{A}(y)$ implies $\pi_{\mathcal{A}}(\mathcal{A}(x)) = \pi_{\mathcal{A}}(\mathcal{A}(y))$. □

6.3 The Universal Property and Information Minimality

Theorem 6.5 (Universal Property of the Admissibility Quotient). *Let $\pi : \mathcal{X} \to \mathcal{M}$ be exact minimal admissibility-preserving. Then there exists a unique continuous injection $\phi : \mathcal{M}_0 \to \mathcal{M}$ with $\pi = \phi \circ \pi_{\mathcal{A}}$.*

Proof. Define $\phi([x]) := \pi(x)$. Since π is exact minimal, $[x] = [y]$ iff $\mathcal{A}(x) = \mathcal{A}(y)$ iff $\pi(x) = \pi(y)$, so ϕ is well-defined and injective. Continuity follows from the quotient topology. Uniqueness follows from surjectivity of $\pi_{\mathcal{A}}$. □

Theorem 6.6 (Information Minimality of the Admissibility Quotient). *Let $C(\mathcal{M}) = I(\mathcal{X}; \mathcal{M})$ denote the mutual information between $\mathcal{X}$ and its projected image. For every exact minimal admissibility-preserving projection $\pi : \mathcal{X} \to \mathcal{M}$:*

$$I(\mathcal{X}; \mathcal{M}) = I(\mathcal{X}; \mathcal{M}_0).$$

Hence the admissibility quotient is the unique information-minimal representation up to isomorphism among exact minimal admissibility-preserving projections.

Proof. By Theorem **??**, $\pi = \phi \circ \pi_{\mathcal{A}}$ with ϕ injective. The data processing inequality gives $I(\mathcal{X}; \mathcal{M}) \leq I(\mathcal{X}; \mathcal{X})$, while the exact minimal condition means π preserves all and only the admissibility-class distinctions. Since ϕ is injective, $I(\mathcal{X}; \mathcal{M}) = I(\mathcal{X}; \mathcal{M}_0)$: neither information is created by ϕ nor is any admissibility-relevant information lost. Any representation with strictly less information would fail to separate some pair of states with distinct admissible futures, violating exact minimality. □

Remark 6.7. The previous draft titled this the "Optimal Reduction" theorem and proved it as a complexity minimization. The information-theoretic formulation is stronger and cleaner. What is established is not that the quotient achieves a minimum among a class of candidates but that all exact minimal projections carry exactly the same information, making the quotient the canonical representative of this unique information class. The theorem is therefore better understood as a uniqueness result than an optimization result.

6.4 Entropy Preservation as Necessary Condition

Proposition 6.8 (Entropy Preservation as Necessary Condition)**.** *In the smooth RSVP setting, if $\pi : \mathcal{X} \to \mathcal{M}$ is admissibility-preserving in the sense of Definition* **??***(1), then*

$$S(x) = S(\pi(x)) \qquad \textit{for all } x \in \mathcal{X}.$$

The converse holds only under the additional assumption that admissible future cones are uniquely determined by their measure within each admissibility class.

Proof. Forward: admissibility preservation implies $\mu(\mathcal{A}(x)) = \mu(\mathcal{A}_{\mathcal{M}}(\pi(x)))$, hence $S(x) = S(\pi(x))$.

Converse failure: two compact sets $A, B \subset \mathcal{X}$ with $\mu(A) = \mu(B)$ but $A \neq B$ furnish a counterexample. A projection mapping x to a point whose induced admissible set is B when $\mathcal{A}(x) = A$ preserves entropy but not admissibility. Such sets exist generically.

Converse under additional assumption: if within each $\sim_{\mathcal{A}}$-class the admissible future cone is the unique compact set realizing its μ-measure, then entropy equality implies set equality. ☐

CHAPTER 7

SALIENCE, NOVELTY, AND THE FAMILIARITY POTENTIAL

7.1 Formal Salience (Metric Setting)

Theorem 7.1 (Salience–Admissibility Equivalence). *Let $D_f : \mathcal{X} \to \mathcal{X}_f$ be the deletion operator removing feature f. In the metric setting, f is salient iff*

$$\mathcal{A}(x) \neq \mathcal{A}(D_f(x)) \qquad \textit{for some } x \in \mathcal{X}.$$

Proof. If f is not salient, removing it does not affect any admissible future cone, so $\mathcal{A}(x) = \mathcal{A}(D_f(x))$ for all x, and f may be discarded without changing any admissibility class. Conversely, if $\mathcal{A}(x) \neq \mathcal{A}(D_f(x))$ for some x, then x and $D_f(x)$ lie in different $\sim_{\mathcal{A}}$-classes. Any exact minimal admissibility-preserving projection must distinguish them, so f cannot be discarded. □

7.2 Differential Salience (Smooth RSVP Setting)

Proposition 7.2 (Salience as Entropy Gradient). *In the smooth RSVP setting, if $\mathcal{A}(x)$ varies smoothly with x and S is differentiable, then the differential criterion*

$$\nabla_f S(x) \neq 0$$

is a computable proxy for salience of feature f at x.

Remark 7.3. This is a proxy, not an equivalence. Zero entropy gradient does not imply non-salience, because equal entropy does not imply equal admissible future sets (Proposition **??**). The set-theoretic criterion of Theorem **??** is primary; the gradient version is available and useful in the smooth setting.

7.3 The Smoothed Familiarity Potential

Definition 7.4 (Smoothed Familiarity Potential). Let $\mathcal{C} \subset \mathcal{X}$ be compact and $\varepsilon > 0$. Define

$$\phi^{\varepsilon}_{\mathcal{C}}(x) = -\log\bigl(\varepsilon + d(x, \mathcal{C})\bigr),$$

where $d(x,\mathcal{C}) = \inf_{c\in\mathcal{C}} d_{\mathcal{X}}(x,c)$. The parameter ε prevents the singularity at corpus points where $d(x,\mathcal{C}) = 0$. The function $\phi^{\varepsilon}_{\mathcal{C}}$ is finite and Lipschitz on all of $\mathcal{X}$ for any $\varepsilon > 0$.

Definition 7.5 (Corpus Congruence and Novelty)**.**

$$\mathrm{CC}^{\varepsilon}(x) = e^{-d(x,\mathcal{C})/(\varepsilon+d(x,\mathcal{C}))}, \qquad \mathrm{Nov}^{\varepsilon}(x) = 1 - \mathrm{CC}^{\varepsilon}(x).$$

Proposition 7.6 (Gradient of the Familiarity Potential)**.** *In the smooth RSVP setting, suppose $d(\,\cdot\,,\mathcal{C})$ is differentiable at x with unique nearest corpus point $c^*(x) = P_{\mathcal{C}}(x)$. Then*

$$\nabla\phi^{\varepsilon}_{\mathcal{C}}(x) = -\frac{\nabla d(x,\mathcal{C})}{\varepsilon + d(x,\mathcal{C})}.$$

On a Riemannian manifold with Euclidean local coordinates at x, where $\nabla d(x,\mathcal{C}) = (x - c^*(x))\,/\,d(x,\mathcal{C})$ (the unit vector toward $c^*(x)$), this reduces to

$$\nabla\phi^{\varepsilon}_{\mathcal{C}}(x) = -\frac{x - c^*(x)}{(\varepsilon + d(x,\mathcal{C}))\,d(x,\mathcal{C})}.$$

The gradient points toward the corpus and defines the inferential flow field $\mathbf{v} = \nabla\phi^{\varepsilon}_{\mathcal{C}}$ in the RSVP framework.

Proof. Apply the chain rule to $\phi^{\varepsilon}_{\mathcal{C}} = -\log(\varepsilon + d(\,\cdot\,,\mathcal{C}))$:

$$\nabla\phi^{\varepsilon}_{\mathcal{C}} = -\frac{1}{\varepsilon + d(x,\mathcal{C})}\,\nabla d(x,\mathcal{C}).$$

On a Riemannian manifold, when $x \notin \mathcal{C}$ and $c^*(x)$ is unique, the gradient of the distance function is the unit-speed geodesic direction from x toward $c^*(x)$, i.e., $\nabla d(x,\mathcal{C}) = \exp_x^{-1}(c^*(x))\,/\,d(x,\mathcal{C})$. In Euclidean local coordinates this is $(x - c^*(x))/d(x,\mathcal{C})$. Substituting yields the stated formula. $\square$

Remark 7.7. The general formula $\nabla\phi^{\varepsilon}_{\mathcal{C}} = -\nabla d(\,\cdot\,,\mathcal{C})\,/\,(\varepsilon + d(\,\cdot\,,\mathcal{C}))$ is valid on any Riemannian manifold wherever the distance function is differentiable. The Euclidean simplification in local coordinates holds when the exponential map at x is approximately isometric, which is guaranteed in a sufficiently small neighborhood of any point by the definition of a Riemannian manifold.

Part III

Anderson′s Epistemology as Theorem

CHAPTER 8

THE PRIORITY OF UNDERSTANDING

Theorem 8.1 (Priority of Understanding)**.** *Let* $(\mathcal{X}, \mathcal{A})$ *be a trajectory space with admissibility operator. Define Understanding as the construction of* $\pi : \mathcal{X} \to \mathcal{M}$*, and Reasoning as an operator* $R : \mathcal{M} \to \mathcal{M}$*. Then:*

1. *Reasoning requires a prior Understanding: without* π*, the domain* $\mathcal{M}$ *of* R *does not exist.*
2. *Understanding does not require Reasoning: the construction of* $\pi_{\mathcal{A}} : \mathcal{X} \to \mathcal{M}_0$ *requires only* $\mathcal{X}$ *and* $\mathcal{A}$*.*

Proof. For (1): R acts on $\mathcal{M}$. $\mathcal{M}$ is the codomain of π. Without a prior π, $\mathcal{M}$ is undefined and R cannot be specified.

For (2): The admissibility quotient $\mathcal{M}_0$ is constructed from $\mathcal{X}$ and $\sim_{\mathcal{A}} \subset \mathcal{X} \times \mathcal{X}$, which is determined by $\mathcal{A}$. No operator on any manifold is presupposed. The manifold $\mathcal{M}_0$ is produced by the construction, not provided as input. □

Remark 8.2. Anderson: "You cannot Reason about that which you do not Understand." Theorem **??** shows this is a structural fact about the logical dependency between projections and operators-on-projections. The proof is almost trivial once the objects are correctly defined — which is itself evidence that Anderson's Priority Thesis is not a contingent empirical claim but a tautology of the correct depth.

CHAPTER 9

THE IDENTIFICATION THEOREM: UNDERSTANDING IS THE QUOTIENT CONSTRUCTION

> Intuitive means that the system can very quickly provide solutions to very complex problems but those solutions may not be correct every time.
>
> *Monica Anderson,* Why AI Works

9.1 The Philosophical Gap

Theorem **??** establishes that Reasoning requires a prior Understanding in the structural sense: a reasoning operator $R : \mathcal{M} \to \mathcal{M}$ presupposes the existence of $\mathcal{M}$, which presupposes the construction of a projection $\pi : \mathcal{X} \to \mathcal{M}$. A mathematician will accept this. A philosopher may respond: the theorem shows that *some* representational construction must precede reasoning; it does not show that Anderson's Understanding is *identical* to the admissibility quotient construction rather than merely a special case of it, or a process that produces the quotient as a byproduct, or something altogether different that has the same structural dependency.

The gap between

$$\text{Understanding} \supseteq \text{construction of admissibility classes}$$

and

$$\text{Understanding} = \text{construction of admissibility classes}$$

is the gap this chapter closes.

The strategy is not definitional. We do not simply declare that Understanding means the quotient construction. Instead we inventory every operation Anderson attributes to Understanding, and prove that each is a property of the admissibility quotient and its associated fields. If Anderson's Understanding does everything the quotient construction does, and nothing the quotient construction does not do, then the identification is established empirically within the framework.

9.2 Anderson's Inventory of Understanding

Anderson attributes the following operations to Understanding across her collected writings, including the 2007 *Artificial Intuition* site and the later *Little Pills* essays. The 2007 site adds the specific mechanism of prediction (understood as admissibility estimation), the nested prediction hierarchy that produces emergent semantics, and the robustness and reliability properties. These are incorporated in the Identification Theorem below.

1. **Salience detection.** Understanding determines what is relevant in an input and what can be discarded.
2. **Corpus congruence.** Understanding measures how similar a new input is to prior experience.
3. **Novelty recognition.** Understanding detects when an input falls outside prior experience.
4. **Abstraction formation.** Understanding moves from low-level data to high-level categories.
5. **Graceful degradation.** Understanding fails smoothly at the edges of competence rather than catastrophically.
6. **Model-free operation.** Understanding operates without a priori symbolic representations of the problem domain.
7. **Corpus-driven improvement.** Understanding improves with more exposure to examples.
8. **Priority over Reasoning.** Understanding is a precondition for Reasoning but not vice versa.

The Identification Theorem establishes that each of these corresponds to a property of the admissibility quotient construction.

9.3 The Identification Theorem

Theorem 9.1 (The Identification Theorem)**.** *In the metric and smooth RSVP settings, the admissibility quotient construction* $\pi_{\mathcal{A}} : \mathcal{X} \to \mathcal{M}_0 = \mathcal{X}/{\sim_{\mathcal{A}}}$ *realizes every operation Anderson attributes to Understanding. Specifically:*

1. ***Salience detection is admissibility sensitivity.*** *The features retained by* $\pi_{\mathcal{A}}$ *are exactly the salient features: those whose deletion changes some admissible future cone.* (*Theorem* **??** *and Corollary* **??**.)
2. ***Corpus congruence is familiarity potential level-set structure.*** *The degree to which* $\pi_{\mathcal{A}}$ *assigns similar representations to similar inputs is measured by* $\mathrm{CC}^{\varepsilon}(x) = e^{-d(x,\mathcal{C})/(\varepsilon+d(x,\mathcal{C}))}$*, the level-set structure of the smoothed familiarity potential* $\phi^{\varepsilon}_{\mathcal{C}}$*.*

3. ***Novelty recognition is departure from the corpus manifold.*** *The system detects novelty as* $\mathrm{Nov}^\varepsilon(x) = 1 - \mathrm{CC}^\varepsilon(x)$*, the distance from the projected input to the learned portion of* $\mathcal{M}_0$*.*

4. ***Abstraction formation is the quotient construction itself.*** *The move from rich trajectory space* $\mathcal{X}$ *to the compressed quotient* $\mathcal{M}_0$ *is exactly the formation of abstract categories from sensory particulars: states sharing admissible futures are collapsed to a single abstract representative.*

5. ***Graceful degradation is continuity of the familiarity potential.*** *Since* $\phi^\varepsilon_\mathcal{C}$ *is Lipschitz on* $\mathcal{X}$*, the system's confidence (measured by* CC^ε*) degrades continuously as inputs move away from the corpus. There is no sharp boundary between competence and incompetence. (Proposition* **??***.)*

6. ***Model-free operation is that*** $\pi_\mathcal{A}$ ***is learned, not hand-authored.*** *The admissibility quotient is determined entirely by* $(\mathcal{X}, \mathcal{A})$*. No a priori symbolic representation of the domain is assumed. CLIO approximates it from data without requiring a human-specified model. (Theorem* **??***.)*

7. ***Corpus-driven improvement is CLIO convergence.*** *As corpus size increases, the empirical CLIO minimizer converges in probability to* $\pi_\mathcal{A}$*, realizing Anderson's observation that continued learning mainly corrects corner-case misunderstandings. (Theorem* **??***.)*

8. ***Priority over Reasoning is structural dependency.*** *The reasoning operator* $R : \mathcal{M}_0 \to \mathcal{M}_0$ *cannot be specified before* $\mathcal{M}_0$ *exists, and* $\mathcal{M}_0$ *is produced by* $\pi_\mathcal{A}$*. (Theorem* **??***.)*

Proof. Each item is a reference to a previously established result. Items (1), (6), (7), and (8) are proved theorems. Items (2) and (3) are consequences of Definition **??** and Proposition **??**. Item (4) is the definition of the quotient: two states are collapsed iff their admissible futures agree, which is precisely the formation of an abstract category from instances. Item (5) follows from the Lipschitz property of $\phi^\varepsilon_\mathcal{C}$: for any $\varepsilon > 0$, $|\phi^\varepsilon_\mathcal{C}(x) - \phi^\varepsilon_\mathcal{C}(y)| \leq \|x - y\|/\varepsilon$, so CC^ε varies continuously. □

9.4 What the Identification Establishes

Theorem **??** does not claim that Anderson had the admissibility quotient in mind. It claims something stronger and more useful: that every operation she attributes to Understanding, described in informal epistemological language developed independently of this framework, turns out to be a property of the admissibility quotient construction that would have been built from the requirements of admissible dynamics alone.

This is the sense in which Anderson's framework is derived rather than merely translated. She arrived at the correct operational characterization of a mathematical object she did not have the tools to construct. The identification is not imposed on her ideas from outside; it is extracted from them by asking what mathematical structure would realize each property she attributes to Understanding.

9.5 What the Identification Does Not Establish

The identification is not a proof that the brain implements the admissibility quotient, or that any specific neural architecture does so. It is a proof that Anderson's informal characterization of Understanding is extensionally equivalent to the quotient construction within this framework. Whether the brain is best modeled by this framework is a separate empirical question that the theorem does not address.

Similarly, the identification does not preclude other mathematical formalizations of Understanding that are also consistent with Anderson's characterization. The admissibility quotient construction is sufficient to realize all of Anderson's attributed properties; whether it is necessary is the subject of Conjecture **??** and further work on the geometry of learning.

CHAPTER 10

THE ANDERSON PROJECTION THEOREM

Theorem 10.1 (The Anderson Projection Theorem)**.** *In the metric setting, let $(\mathcal{X}, \mathcal{A})$ be a trajectory space with Hausdorff-continuous admissibility operator. The following are equivalent.*

1. *A projection $\pi : \mathcal{X} \to \mathcal{M}$ performs Anderson's Epistemic Reduction: it retains exactly those distinctions between states relevant to admissible futures and discards all others.*
2. *π is exact minimal admissibility-preserving: $\pi(x) = \pi(y)$ iff $\mathcal{A}(x) = \mathcal{A}(y)$.*
3. *π factors through the admissibility quotient as a homeomorphism: $\pi = \phi \circ \pi_{\mathcal{A}}$ for some homeomorphism $\phi : \mathcal{M}_0 \to \mathcal{M}$.*
4. *The features discarded by π are exactly the non-salient features: those whose deletion leaves every admissible future cone unchanged.*

Moreover, in the smooth RSVP setting, each of (1)–(4) implies entropy preservation: $S(x) = S(\pi(x))$ for all x. Entropy preservation alone does not imply any of (1)–(4).

Proof. (1) $\Leftrightarrow$ (4): Anderson's Reduction retains the salient and discards the non-salient. By Theorem **??**, non-salient features are exactly those whose deletion leaves every admissible future cone unchanged. So (1) holds iff π discards exactly those features, which is (4).

(2) $\Leftrightarrow$ (4): Exact minimal admissibility-preservation means π collapses x and y iff $\mathcal{A}(x) = \mathcal{A}(y)$. This collapses exactly the distinctions between states with identical admissible futures, retaining exactly the salience-relevant distinctions, which is (4).

(2) $\Rightarrow$ (3): By Theorem **??**, π factors through $\pi_{\mathcal{A}}$ via an injection ϕ. Since π is surjective, ϕ is bijective, hence a homeomorphism by compactness of $\mathcal{M}_0$.

(3) $\Rightarrow$ (2): If $\pi = \phi \circ \pi_{\mathcal{A}}$ with ϕ a homeomorphism, then $\pi(x) = \pi(y)$ iff $\pi_{\mathcal{A}}(x) = \pi_{\mathcal{A}}(y)$ iff $\mathcal{A}(x) = \mathcal{A}(y)$.

Entropy preservation: follows from Proposition **??**, since exact minimal admissibility-preservation implies admissibility preservation in sense (1) of Definition **??**. Converse failure is the content of Proposition **??**. □

Corollary 10.2 (Salience Criterion Derived)**.** *Anderson's instruction that Reduction retains the salient and discards the non-salient is equivalent to requiring that Reduction is an exact minimal admissibility-preserving projection.*

Remark 10.3. A feature matters in Anderson's sense precisely because it bends the admissible future space. Salience is not psychological; it is geometric. The theorem says that Epistemic Reduction, admissibility-preservation, factoring through the quotient, and salience-filtering are four characterizations of the same mathematical object.

CHAPTER 11

THE CLIO OPERATOR AND CORPUS-DRIVEN LEARNING

11.1 The CLIO Functional

Definition 11.1 (CLIO Reduction Functional)**.** The CLIO reduction functional is

$$L_{\mathrm{CLIO}}[\pi] = L_{\mathsf{task}}(\pi) + \lambda\,\mathbb{E}_{x\sim\mathcal{C}}\big[d_H\big(\mathcal{A}(x),\,\mathcal{A}_{\mathcal{M}}(\pi(x))\big)\big] + \mu\,I(\mathcal{X};\mathcal{M}),$$

where L_{task} is task-specific performance loss, the second term is the empirical admissibility loss, and $\lambda, \mu > 0$ are regularization parameters.

Theorem 11.2 (CLIO Consistency, Conditional Form)**.** *Assume:*

- (*C1*) *The hypothesis class* Π *of candidate projections is compact in a suitable function space topology.*
- (*C2*) *The CLIO loss satisfies uniform convergence over* Π*: empirical loss converges uniformly to population loss as* $|\mathcal{C}_n| \to \infty$*.*
- (*C3*) *The admissibility quotient projection* $\pi_{\mathcal{A}}$ *is identifiable as the unique population minimizer of* L_{CLIO} *within* Π*.*

Then $\pi_n^* := \arg\min_{\pi\in\Pi} L_{\mathrm{CLIO}}[\pi]$ *over* $\mathcal{C}_n$ *converges in probability to* $\pi_{\mathcal{A}}$ *as* $n \to \infty$*.*

Proof. Under (C1) and (C2), any sequence of empirical minimizers has population risk converging to the minimum (standard M-estimation). Under (C3), the unique population minimizer is $\pi_{\mathcal{A}}$, so convergence in probability follows by the well-specification argument. □

Remark 11.3. Conditions (C1)–(C3) are non-trivial. Condition (C1) is a model complexity constraint. Condition (C2) is a uniform law of large numbers that holds under standard covering number bounds. Condition (C3) requires that CLIO uniquely identifies the admissibility quotient, which depends on adequate corpus coverage of $\mathcal{X}$. These conditions cannot be assumed automatically and must be verified application by application.

CHAPTER 12

ADMISSIBILITY-COMPATIBLE LEARNING AND LOCAL EXTENSION STABILITY

12.1 Formalization of Learning

The admissibility quotient is initially constructed from a fixed $\mathcal{X}$ and $\mathcal{A}$. Learning corresponds to incorporating a novel trajectory $u \in \mathcal{X}$ into the corpus, which updates the projection to cover the new equivalence class $[u]$.

Definition 12.1 (Learning Extension Operator)**.** Let $\mathcal{M}_0 = \mathcal{X}/{\sim_{\mathcal{A}}}$ be the admissibility quotient and $u \in \mathcal{X}$ a novel input not previously represented in the corpus. Define the learning extension operator

$$L_u : \mathcal{M}_0 \longrightarrow \mathcal{M}_0',$$

where $\mathcal{M}_0' = (\mathcal{X} \cup \{u\})/{\sim_{\mathcal{A}}}'$ is the admissibility quotient after incorporating u, and ${\sim_{\mathcal{A}}}'$ is the induced equivalence relation on the extended trajectory space.

Definition 12.2 (Admissibility-Compatible Learning)**.** The learning extension L_u is *admissibility-compatible* if $\mathcal{M}_0'$ is diffeomorphic to a smooth extension of $\mathcal{M}_0$ as a Riemannian manifold. That is, there exists an open embedding $\iota : \mathcal{M}_0 \hookrightarrow \mathcal{M}_0'$ such that $L_u \circ \iota$ is the identity on the original quotient, and the smooth structure of $\mathcal{M}_0'$ restricts to that of $\mathcal{M}_0$ on $\iota(\mathcal{M}_0)$.

12.2 Local Extension Stability

Theorem 12.3 (Local Extension Stability)**.** *In the smooth RSVP setting, suppose $S \in C^2(\mathcal{M}_0)$ and let $m_u = \pi_{\mathcal{A}}(u) \in \mathcal{M}_0'$ be the projected location of the novel input. If*

$$\sup_{m \in U} \|\nabla^2 S(m)\| < K_{\max}$$

for some open neighborhood U of m_u in $\mathcal{M}_0'$, then the learning extension L_u produces a locally stable deformation of admissibility structure: for all h in the tangent space at m_u,

$$\big|S(m_u + h) - S(m_u)\big| \;\le\; \|\nabla S(m_u)\|\,\|h\| + \tfrac{1}{2}\,K_{\max}\,\|h\|^2.$$

In particular, the admissibility geometry varies continuously near m_u and the learning extension introduces no singularity in S.

Proof. The bound follows directly from Taylor's theorem applied to $S \in C^2$ at m_u:

$$S(m_u + h) = S(m_u) + \nabla S(m_u) \cdot h + \tfrac{1}{2}\, h^T (\nabla^2 S)(\xi)\, h$$

for some ξ on the segment from m_u to $m_u + h$, which lies in U for $\|h\|$ sufficiently small. The operator norm bound $\|(\nabla^2 S)(\xi)\| < K_{\text{max}}$ gives $|h^T(\nabla^2 S)(\xi)h| \le K_{\text{max}}\|h\|^2$, and the stated inequality follows by the triangle inequality. □

Remark 12.4. Theorem **??** is a genuine one-directional result: bounded curvature near m_u implies locally stable deformation. The converse — that locally stable deformation requires bounded curvature — is not proved here and is expected to be false in full generality, since other geometric quantities (torsion, higher-order terms) may compensate for large Hessian in special cases.

Conjecture 12.5 (Curvature–Admissibility Threshold)**.** *Admissibility-compatible learning in the sense of Definition* **??** *holds if and only if the Hessian of* S *is bounded by a threshold* K_{max} *determined by the global geometry of* $\mathcal{M}_0$.

The missing ingredient for a proof in the forward direction is a result showing that unbounded Hessian necessarily produces a singularity (rather than merely large but finite deformation) in the admissibility structure. The missing ingredient for the converse is a characterization of the conditions under which large Hessian is avoidable through reparametrization.

Corollary 12.6 (Winston's Condition, Conditional)**.** *Under the hypotheses of Theorem* **??***, Patrick Winston's aphorism "you can only learn what you almost already know" follows in the following sense: learning extension is locally stable when the novel input projects to a region of* $\mathcal{M}_0$ *with bounded curvature in* S*, i.e., when the new equivalence class is geometrically accessible from the existing manifold structure.*

CHAPTER 13

WHY DEEP LEARNING WORKS: APPROXIMATE ADMISSIBILITY REDUCTION

Deep Learning Performs Epistemic Reduction.

Monica Anderson, Why Deep Learning Works

13.1 Anderson's Claim

Anderson argued repeatedly that the success of deep learning after 2012 is explained by the fact that deep neural networks perform Epistemic Reduction autonomously, without requiring a human agent to first understand and encode the domain. She further argued that the reason deep networks must be deep is that Reduction must be performed hierarchically: a feature can only be identified as salient or non-salient at the level of abstraction where its relevance becomes apparent, and this level is not available until lower levels have already performed their reductions.

This chapter shows that these claims follow as theorems from the framework developed in Parts II and III, given a standard account of what deep learning does computationally.

13.2 Deep Learning as Sequential Partial Quotient

Definition 13.1 (Layer-wise Projection)**.** A deep neural network with L layers computes a composition of projections

$$\mathcal{X} = \mathcal{X}^{(0)} \xrightarrow{\pi^{(1)}} \mathcal{X}^{(1)} \xrightarrow{\pi^{(2)}} \cdots \xrightarrow{\pi^{(L)}} \mathcal{X}^{(L)},$$

where each $\pi^{(\ell)} : \mathcal{X}^{(\ell-1)} \to \mathcal{X}^{(\ell)}$ is a parametric map (a layer) and the composition $\pi = \pi^{(L)} \circ \cdots \circ \pi^{(1)}$ is the full network projection.

Proposition 13.2 (Layered Reduction as Sequential Partial Quotient)**.** *Each layer $\pi^{(\ell)}$ performs a partial admissibility quotient: it maps states that are indistinguishable at abstraction level ℓ (given the representations available at level $\ell - 1$) to the same* *output, while preserving distinctions that are admissibility-relevant at that level. The full network projection approximates the global admissibility quotient* $\pi_A : \mathcal{X} \to \mathcal{M}_0$.

Remark 13.3. Proposition **??** is an interpretive claim, not a theorem from the preceding machinery. It identifies the computational function of each layer with a step in the admissibility quotient hierarchy of Theorem **??**. The precise conditions under which a trained layer implements a partial admissibility quotient depend on the training objective, architecture, and data distribution, and are not derived here. The proposition frames the relationship between deep learning and the quotient hierarchy; the CLIO Convergence Theorem (Theorem **??**) makes the convergence claim precise under stated conditions.

13.3 Why Deep Learning Must Be Deep: A Derived Theorem

Theorem 13.4 (Depth is Necessary for Multi-Level Salience)**.** *In the metric setting, suppose the admissibility quotient hierarchy* $\mathcal{X}_0 \to \mathcal{X}_1 \to \cdots \to \mathcal{X}_N$ *has depth* $N > 1$*: the sequence does not stabilize after one step. Then no single-layer projection* $\pi : \mathcal{X} \to \mathcal{M}$ *can simultaneously achieve the salience-filtering performed by the full hierarchy.*

Proof. By Theorem **??**, the hierarchy satisfies $I(\mathcal{X}_{n+1}) \leq I(\mathcal{X}_n)$ with equality iff the $\sim_{\mathcal{A}_n}$-classes are all singletons (no compression at level n). If the hierarchy has depth $N > 1$, there exist features f that are salient at level $k > 1$ but not at level 0: features whose admissibility-sensitivity is only revealed after lower-level reductions have been applied.

A single-layer projection $\pi : \mathcal{X} \to \mathcal{M}$ operates at level 0. By Theorem **??**, any feature discarded by π cannot become salient through operations internal to $\mathcal{M}$. Features that are non-salient at level 0 but salient at level $k > 1$ would therefore be discarded by π and could never be identified. The single-layer projection therefore cannot achieve the salience-filtering that requires recognition at level k. □

Remark 13.5. Theorem **??** is the formal version of Anderson's claim that "Deep Learning is deep because you can only do Reduction by discarding the irrelevant if you Understand what is relevant and irrelevant at each different level of Abstraction." The theorem shows this is not an architectural preference but a logical necessity: when the admissibility quotient hierarchy has non-trivial depth, a single compression step cannot replicate its filtering. Depth is structurally required.

13.4 Deep Learning Converges to the Admissibility Quotient

The central synthesis theorem combines the CLIO Convergence result with the layer-wise structure of deep networks.

Theorem 13.6 (Deep Learning as Approximate Admissibility Reduction)**.** *Under the conditions of Theorem **??** and the layer-wise architecture of Definition **??**, the full network projection π_n trained on corpus $\mathcal{C}_n$ satisfies:*

$$\pi_n \xrightarrow{p} \pi_{\mathcal{A}} \qquad \text{as } |\mathcal{C}_n| \to \infty,$$

where $\pi_{\mathcal{A}}$ is the admissibility quotient projection. Deep learning is therefore a numerical approximation to the admissibility quotient, carried out by hierarchical partial reduction across layers.

Proof. The full network projection π_n is the composition $\pi_n^{(L)} \circ \cdots \circ \pi_n^{(1)}$. Under the conditions (C1)–(C3) of Theorem **??**, the empirical CLIO minimizer over the corpus $\mathcal{C}_n$ converges in probability to $\pi_{\mathcal{A}}$. Since π_n is the CLIO minimizer within the layer-wise hypothesis class, the convergence follows from Theorem **??** applied to this class. □

Remark 13.7. Theorem **??** is the most important bridge in the monograph between Anderson's informal epistemology and contemporary AI. Anderson argued that deep learning succeeds because it performs Epistemic Reduction. The theorem makes this precise: under CLIO consistency conditions, the learned network projection converges to the admissibility quotient — the unique information-minimal admissibility-preserving representation of the domain.

Deep learning succeeds not because it finds a good heuristic for the task but because, given sufficient data and an adequate architecture, it approximates the canonical compression that the geometry of the domain demands. The data processing inequality ensures this cannot be beaten by any fixed-architecture system; the CLIO functional ensures that the network is rewarded for admissibility preservation; and the convergence theorem ensures that the approximation improves with data.

Anderson's claim — that understanding is what deep learning provides and symbolic AI cannot — is now a theorem: deep learning converges to the admissibility quotient, which is the exact mathematical object that Anderson's Understanding requires.

13.5 What Deep Learning Cannot Do

The convergence result has a natural boundary. Deep learning approximates the admissibility quotient of the domain from which the corpus is drawn. It does not construct an admissibility quotient for problems outside this domain, for problems requiring extrapolation beyond the corpus manifold, or for problems whose admissibility structure changes faster than the corpus can track.

Corollary 13.8 (Generalization Boundary)**.** *Under the conditions of Theorem **??**, the learned projection π_n converges to $\pi_{\mathcal{A}}$ within the corpus-covered region of $\mathcal{X}$. For inputs x with $\mathrm{Nov}^{\varepsilon}(x)$ close to one (far from the corpus manifold), the projection error $d_H(\mathcal{A}(x), \mathcal{A}_{\mathcal{M}}(\pi_n(x)))$ may be large. The system's behavior in this region is governed by the smoothed familiarity potential $\phi^{\varepsilon}_{\mathcal{C}}$: it will generalize in the direction of the nearest corpus region, but with increasing error as corpus distance grows.*

Remark 13.9. Corollary **??** formalizes why large language models and deep networks can perform impressively within their training distribution while failing unexpectedly on inputs that are semantically novel. The failure is not a flaw in deep learning per se; it is a direct consequence of what deep learning is: an approximation to the admissibility quotient of the training distribution. Inputs outside that distribution are handled by the familiarity potential, which provides a continuous signal of distance from the learned manifold but makes no guarantees about admissibility preservation at large distances.

Chapter 14

Brittleness, Projection Supremacy, and Holistic Robustness

Definition 14.1 (Projection Supremacy)**.** A cognitive system exhibits Projection Supremacy if it applies $R : \mathcal{M} \to \mathcal{M}$ in contexts where $d_H(\mathcal{A}(x), \mathcal{A}_{\mathcal{M}}(\pi(x))) > 0$ without detecting or compensating for this error.

Theorem 14.2 (Brittleness from Projection Supremacy)**.** *A system exhibiting Projection Supremacy produces unmodulated errors for inputs x with $d_H(\mathcal{A}(x), \mathcal{A}_{\mathcal{M}}(\pi(x))) > 0$, because it has no internal signal for confidence degradation.*

Proof. Where projection error is zero, the system's responses are calibrated to the correct admissible future set. Where error is positive, the system applies R to an incorrect admissible future representation. Projection Supremacy means the system has no variable representing the discrepancy; therefore no mechanism modulates its confidence or output distribution. Errors are unmodulated by the degree of projection failure. □

Remark 14.3. This formalizes Anderson's observation that classical AI systems work in the laboratory but fail spectacularly at the edges of competence: the hand-crafted model is treated as complete, so graceful degradation is structurally impossible.

Proposition 14.4 (Holistic Graceful Degradation)**.** *A Holistic system with smoothed familiarity potential ϕ_C^ε degrades gracefully at the edges of competence: for inputs with large $\mathrm{Nov}^\varepsilon(x)$, the gradient $\nabla\phi_C^\varepsilon(x)$ provides a continuous signal of distance from the corpus, enabling modulation of the output distribution.*

14.1 Interpretive Attractors and Epistemic Identity

A recurring feature of cognition is that understanding rarely exists as an isolated collection of propositions. Instead, it tends to organize itself into coherent interpretive frameworks through which new experiences are filtered and evaluated. This phenomenon has a natural formalization in admissibility geometry.

Definition 14.5 (Interpretive Attractor)**.** An *interpretive attractor* is a compact region $F \subseteq \mathcal{M}_0$ of the admissibility quotient satisfying two conditions:

1. **Internal accessibility:** For $m, m' \in F$, the transition $m \to m'$ lies in $\mathcal{A}_{\mathcal{M}_0}(m)$, so movement within F is admissible.

2. **Boundary asymmetry:** Transitions from F to $\mathcal{M}_0 \setminus F$ have significantly higher cost (in terms of admissibility loss or familiarity potential) than transitions within F:

$$\inf_{m \in F,\, m' \notin F} d_H\big(\mathcal{A}_{\mathcal{M}_0}(m),\, \{m'\}\big) \gg \sup_{m, m' \in F} d_H\big(\mathcal{A}_{\mathcal{M}_0}(m),\, \{m'\}\big).$$

An interpretive framework is a cognitive system whose learned projection π_n maps a large fraction of the corpus into a single interpretive attractor F.

Proposition 14.6 (Interpretive Frameworks as Self-Reinforcing Projections)**.** *In the smooth RSVP setting, if a system's projection π_n is concentrated in an interpretive attractor F, then new inputs u with $\pi_n(u) \in F$ increase $\mathrm{CC}^{\varepsilon}(u)$ and reinforce the existing manifold structure, while inputs u with $\pi_n(u) \notin F$ generate a gradient $\nabla \phi_{\mathcal{C}}^{\varepsilon}(u)$ pointing toward F rather than toward the true corpus region. The attractor therefore becomes self-reinforcing: new experience is assimilated into the existing framework rather than expanding the projection.*

Remark 14.7. Proposition **??** does not imply irrationality. It is a natural consequence of epistemic reduction. Any system that compresses experience into a tractable representation must privilege certain distinctions over others. Once a projection has been learned, new observations are interpreted through that projection. Understanding therefore produces not merely representations of reality but stable *modes of seeing* reality. These modes are what Anderson's epistemology means by frameworks: not theories in the formal sense but attractors in the admissibility manifold that organize incoming experience before it reaches the level of explicit reasoning.

14.2 Communities as Shared Corpora

Corpus Congruence is typically discussed at the level of individual systems. But many of the most powerful forms of understanding arise collectively. The admissibility framework extends naturally to this case.

Definition 14.8 (Distributed Corpus)**.** A *distributed corpus* is a collection $\{\mathcal{C}_i\}_{i \in I}$ of individual corpora maintained by agents $i \in I$, together with an exchange process $\mathcal{E} : \mathcal{C}_i \times \mathcal{C}_j \to \mathcal{C}_i'$ by which agents modify their corpora through interaction with others. The *community corpus* is the aggregate $\mathcal{C} = \bigcup_i \mathcal{C}_i$.

Proposition 14.9 (Community Semantic Manifold)**.** *Under repeated exchange, agents in a community converge toward a shared admissibility quotient $\mathcal{M}_{\mathrm{comm}} = \mathcal{C}/\!\sim_{AC}$, the admissibility quotient of the community corpus. Agents who share more exchange time develop closer familiarity potentials $\phi_{\mathcal{C}i}^{\varepsilon}$ and therefore more closely aligned salience structures.*

Remark 14.10. This observation explains why individuals who encounter identical external events may arrive at radically different conclusions. The difference need not arise from logical disagreement. It arises because the events are projected through different familiarity potentials $\phi^{\varepsilon}_{\mathcal{C}i}$, placing them at different locations within distinct community manifolds. An event that is highly familiar within one community's quotient manifold may be novel within another's, producing different salience evaluations, different interpretation bundles, and therefore different responses.

Understanding is therefore not merely personal. It is a collective geometric structure maintained through continual interaction among many learners. The community corpus is the aggregate object; the community semantic manifold is its admissibility quotient; and social communication is the exchange process that keeps the distributed corpus coherent over time.

14.3 Projection Capture and Epistemic Rigidity

The success of a projection can become the source of its own failure. This section formalizes a phenomenon that extends Projection Supremacy from a static condition to a dynamic one.

Definition 14.11 (Projection Capture)**.** A system exhibits *projection capture* when the learned projection π_n has become insulated from revision: the CLIO update gradient $\nabla L_{\mathrm{CLIO}}[\pi_n]$ is suppressed to near zero by the system's own confidence structure, so that even inputs with large $\mathrm{Nov}^{\varepsilon}(x)$ do not produce meaningful updates to π_n.

Proposition 14.12 (Projection Capture as Dynamic Projection Supremacy)**.** *Projection capture is the dynamic analogue of Projection Supremacy. Projection Supremacy (Definition **??**) is a static condition: the system applies R without detecting projection error. Projection capture is a process condition: the system has lost the ability to* update *π even when projection error is present, because the familiarity signal $\mathrm{CC}^{\varepsilon}(x)$ has been saturated by prior reinforcement within the interpretive attractor.*

Proof. By Definition **??**, the CLIO update gradient is suppressed. This means that inputs with positive projection error $d_H(\mathcal{A}(x), \mathcal{A}_{\mathcal{M}}(\pi(x))) > 0$ fail to drive π toward $\pi_{\mathcal{A}}$. The system therefore persists in a state of Projection Supremacy across all inputs in the attractor's basin of influence, satisfying Definition **??** dynamically rather than by initial construction. □

Remark 14.13. Projection capture differs from ordinary error. An incorrect belief can be corrected within an existing manifold by operations $R : \mathcal{M} \to \mathcal{M}$. Projection capture occurs when the manifold itself becomes insulated from revision: new admissibility structures cannot become salient because they are discarded by a projection that has become self-sealing. By Theorem **??** (Feyerabend–Anderson Constraint), features discarded by a fixed projection cannot be recovered through internal operations. Projection capture is precisely the condition where a formerly learned projection has become, in functional terms, indistinguishable from a fixed one.

The health of an epistemic system therefore depends not only on the quality of its current projection but on its capacity to revise that projection when new admissibility structures become visible. This is the formal content of the Admissible Discovery Principle (Definition **??**): the projection must remain open to revision until the data is sufficient to determine which features are genuinely admissibility-sensitive. Projection capture is the violation of this principle by a system whose prior successes have insulated it from the data it most needs.

CHAPTER 15

CORPUS CONGRUENCE AS POTENTIAL FIELD

Proposition 15.1 (Potential Subsumes Similarity). $\mathrm{CC}^{\varepsilon}(x)$ *is monotonically related to distance from* $\mathcal{C}$*: large when* x *is near* $\mathcal{C}$*, small when far. The level sets of* $\phi_{\mathcal{C}}^{\varepsilon}$ *recover Anderson's similarity ordering; the gradient* $\nabla\phi_{\mathcal{C}}^{\varepsilon}$ *adds the directional structure needed for inference as a dynamical process.*

Remark 15.2. Anderson's original Corpus Congruence is a similarity measure. The potential field formulation enriches this: the corpus defines a potential landscape over $\mathcal{X}$, and $\nabla\phi_{\mathcal{C}}^{\varepsilon}$ is the RSVP inferential flow field $\mathbf{v}$. The system does not merely compare inputs against stored examples; it moves through a semantic potential landscape generated by the corpus. The level-set structure recovers Anderson's ordering, and the gradient structure models inference as a dynamical process.

Part IV

Extensions and Applications

Chapter 16

Artificial Intuition as Learned Projection

Intuition operates on events, not theories.

Monica Anderson, Artificial Intuition, *2007*

16.1 The 2007 Framework and Its Central Claim

Anderson's 2007 *Artificial Intuition* website introduced the core epistemological programme that her later *Little Pills* essays refined. Its central claim is compressed in the sentence that opens this chapter: intuition operates on events, not theories.

Read literally, the claim is incomplete. A system that operates on events must have some internal structure to do so usefully. Anderson acknowledges this elsewhere, discussing world models, semantics, and concept hierarchies. The apparent contradiction dissolves with the correct gloss: Intuition is *pre-theoretical* rather than theory-free. A logical system begins with an explicit symbolic model and derives consequences. An intuitive system begins with admissibility structure — correlations, precedents, accessible future cones — and only later develops abstractions from that structure. The representational manifold appears at the end, not the beginning. This is precisely the ordering captured by the admissibility quotient construction and formalized as the Priority Theorem (Theorem **??**).

16.2 Artificial Intuition Defined

Definition 16.1 (Artificial Intuition)**.** Artificial Intuition is the process of approximating the admissibility quotient projection $\pi_{\mathcal{A}} : \mathcal{X} \to \mathcal{M}_0$ from experience alone, without first specifying a symbolic theory of the domain:

$$\text{Artificial Intuition} = \text{approximation of } \pi_{\mathcal{A}} : \mathcal{X} \to \mathcal{M}_0 \text{ from corpus } \mathcal{C} \subset \mathcal{X}.$$

No symbolic theory appears in the construction.

Remark 16.2. Definition **??** does not define the cognitive phenomenon of human intuition. It defines what Anderson's programme was constructing: a system that learns from events, produces semantics and prediction, and does so without explicit models. Each of those informal properties is a consequence of corpus-driven admissibility quotient approximation.

16.3 Intuitive States and Admissibility Neighborhoods

Definition 16.3 (Intuitive State)**.** An *intuitive state* at $x \in \mathcal{X}$ is a probability measure $I_x \in \mathcal{P}(\mathcal{X})$ supported primarily on the admissibility neighborhood of x:

$$\operatorname{supp}(I_x) \;\subseteq\; \bigl\{y \in \mathcal{X} : d_H\bigl(\mathcal{A}(x), \mathcal{A}(y)\bigr) < \varepsilon\bigr\}.$$

An intuitive state stores neighborhoods in admissibility geometry, not rules or propositions.

Remark 16.4. Anderson writes that intuition "tracks events" and stores "correlations between preceding and consequent events." In admissibility geometry, a correlation between events e and e' is the statement that $e' \in \mathcal{A}(e)$, or more generally that $d_{\mathcal{X}}(e', \mathcal{A}(e))$ is small. An intuitive state is a probability measure over the admissibility cone, weighted by prior frequency of occurrence in the corpus.

16.4 Prediction as Admissibility Estimation

Anderson's 2007 site opens by arguing that "the purpose of Intelligence is Prediction." This is often read as point forecasting of the next state. The admissibility reinterpretation is more precise and more powerful.

Definition 16.5 (Prediction in the Admissibility Framework)**.** A predictor $\hat{\mathcal{A}} : \mathcal{X} \to 2^{\mathcal{X}}$ approximates the true admissibility operator $\mathcal{A}$. Prediction quality is measured by:

$$L_P = \mathbb{E}_x\bigl[d_H\bigl(\hat{\mathcal{A}}(x), \mathcal{A}(x)\bigr)\bigr].$$

Intelligence is the ability to maintain a low-error estimate of $\mathcal{A}$: to minimize L_P.

Remark 16.6. This reinterpretation explains several features of Anderson's discussion. Why does she emphasize ambiguity? Because multiple admissible futures must be tracked, not just the most likely one. Why robustness? Because the admissibility cone must be estimated from incomplete observations. Why does prediction connect to semantics? Because higher-level semantic categories are exactly the equivalence classes of the admissibility quotient: states sharing the same accessible future structure.

16.5 The Nested Prediction Hierarchy and Emergent Semantics

Anderson argues that semantics emerges from nested predictions: lower-level predictions are predicted by higher-level predictors, and this cascading structure produces semantic categories. In the admissibility framework this is the recursive quotient hierarchy of Theorem **??**.

Definition 16.7 (Nested Prediction Hierarchy)**.** Define $\mathcal{X}_0 = \mathcal{X}$ and $\mathcal{A}_0 = \mathcal{A}$. Recursively:

$$\mathcal{X}_{k+1} = \mathcal{X}_k/{\sim_{\mathcal{A}k}}, \qquad \mathcal{A}_{k+1} : \mathcal{X}_{k+1} \to 2^{\mathcal{X}_{k+1}} \text{ induced.}$$

The sequence $(\mathcal{X}_k, \mathcal{A}_k)_{k\geq 0}$ is the nested prediction hierarchy; semantics emerges as repeated quotienting.

Proposition 16.8 (Emergent Semantics as Admissibility Attractors)**.** *A semantic category at level k is a stable region $S \subseteq \mathcal{X}_k$ satisfying:*

$$\forall x, y \in S : \quad d_H\big(\mathcal{A}_k(x), \mathcal{A}_k(y)\big) < \varepsilon.$$

Semantic categories are not primitives. They are admissibility attractors: regions of the quotient manifold where the accessible future structure is approximately constant. Words, concepts, objects, social roles, and abstractions are all admissibility attractors at appropriate levels.

Remark 16.9. Anderson writes that meaning "is an emergent property of the sentence, the paragraph, the rules of language, the topic under discussion, and the shared world model of the writer and the reader." Meaning is not located in components but in the admissibility structure that spans them. The shared world model is the shared admissibility operator; semantics is agreement on accessible futures.

16.6 The Theory-Free Learning Theorem

Theorem 16.10 (Theory-Free Learning Theorem)**.** *Under the conditions of Theorem* **??***, empirical projections π_n learned from a growing corpus $\mathcal{C}_n$ satisfy:*

$$\pi_n \xrightarrow{p} \pi_{\mathcal{A}} \qquad \text{as } |\mathcal{C}_n| \to \infty,$$

without specifying any symbolic theory of the domain.

Proof. This is Theorem **??** under the observation that L_{CLIO} requires no symbolic domain theory: only a corpus $\mathcal{C}_n \subset \mathcal{X}$, a task loss, and empirical admissibility statistics. No grammar, ontology, or rule system appears as input. □

Remark 16.11. Theorem **??** formalizes Anderson's claim that "Intuition learns directly from events." The CLIO operator produces an approximation to the admissibility quotient from event co-occurrence statistics alone. There is no theory-in, quotient-out step; the quotient emerges from the data. This is the mathematical content of Anderson's claim that Intuition is pre-theoretical: the representational manifold is produced by the learning process, not provided as prior knowledge.

16.7 Prediction Generates the Quotient Hierarchy

The deepest synthesis between Anderson's 2007 Artificial Intuition programme and the RSVP framework is the following theorem: prediction itself generates the admissibility quotient hierarchy. Understanding is not an independent faculty that happens to resemble prediction; it *is* the stable equivalence structure that emerges from repeated prediction.

Definition 16.12 (Prediction Profile)**.** The *prediction profile* of a state $x \in \mathcal{X}$ with respect to a predictor $\hat{\mathcal{A}} : \mathcal{X} \to 2^{\mathcal{X}}$ is the predicted accessible future cone $\hat{\mathcal{A}}(x)$. Two states $x, y \in \mathcal{X}$ are *prediction-equivalent* if their prediction profiles converge asymptotically as prediction improves:

$$x \ \sim_P \ y \quad \Longleftrightarrow \quad \lim_{n \to \infty} d_H\big(\hat{\mathcal{A}}_n(x), \hat{\mathcal{A}}_n(y)\big) = 0,$$

where $\hat{\mathcal{A}}_n$ is the predictor trained on corpus $\mathcal{C}_n$.

Theorem 16.13 (Semantic Emergence by Prediction Quotients)**.** *Under the conditions of Theorem* **??***, the prediction-equivalence relation* $\sim_P$ *converges in probability to the admissibility equivalence* $\sim_{\mathcal{A}}$ *as* $|\mathcal{C}_n| \to \infty$*. Consequently:*

1. *The prediction quotient* $\mathcal{X}/\sim_P$ *converges to the admissibility quotient* $\mathcal{M}_0 = \mathcal{X}/\sim_{\mathcal{A}}$*.*
2. *Every stable semantic category — a region of* $\mathcal{M}_0$ *where the accessible future structure is approximately constant (Proposition* **??***) — is an attractor of the prediction-equivalence dynamics.*
3. *Semantics = Stable Equivalence Classes of Prediction.*

Proof. By Theorem **??**, $\hat{\mathcal{A}}_n(x) \to \mathcal{A}(x)$ in probability for all x as $|\mathcal{C}_n| \to \infty$ (under conditions C1–C3). Therefore $d_H(\hat{\mathcal{A}}_n(x), \hat{\mathcal{A}}_n(y)) \to d_H(\mathcal{A}(x), \mathcal{A}(y))$ in probability. The prediction-equivalence $x \sim_P y$ holds (in the limit) iff $d_H(\mathcal{A}(x), \mathcal{A}(y)) = 0$, which is exactly $x \sim_{\mathcal{A}} y$ (Definition **??**). Thus $\sim_P \to \sim_{\mathcal{A}}$ in probability, and the prediction quotient converges to $\mathcal{M}_0$. Item (2) follows from Proposition **??**: stable regions of $\mathcal{M}_0$ are exactly those where $\mathcal{A}$ is approximately constant, which is where prediction profiles stabilize. Item (3) restates items (1) and (2) in Anderson's language. □

Remark 16.14. Theorem **??** establishes the deepest link in the monograph between Anderson's epistemology and the RSVP framework. Anderson's claim that semantics emerges from nested prediction is not merely analogically connected to the admissibility quotient hierarchy: it is mathematically equivalent to it, under CLIO consistency conditions. The full implication chain is now derivable within the framework:

$$\text{Prediction} \implies \text{Semantics} \implies \text{Understanding} \implies \text{Reasoning}.$$

The first arrow is Theorem **??**: stable prediction equivalence classes are the semantic manifold $\mathcal{M}_0$. The second arrow is the Identification Theorem (Theorem **??**): the admissibility quotient construction realizes every operation Anderson attributes to Understanding. The third arrow is the Priority Theorem (Theorem **??**): Reasoning is an operator on $\mathcal{M}_0$, which Understanding constructs. Anderson proposed this ordering in 2007. The present framework proves it.

16.8 Ambiguity, Interpretation Bundles, and Premature Collapse

Definition 16.15 (Interpretation Bundle)**.** The *interpretation bundle* at observation x is a probability measure over the quotient manifold:

$$\rho_x = \sum_i w_i \, \delta_{m_i}, \quad m_i \in \mathcal{M}_0, \quad \sum_i w_i = 1, \quad w_i \geq 0.$$

Each $m_i = [x_i]$ is an admissibility class compatible with the observation; w_i is its posterior probability given the evidence. Interpretation collapse occurs when $H(\rho_x) < H_{\text{min}}$; until then, the full distribution is maintained.

Proposition 16.16 (Premature Collapse as Projection Supremacy)**.** *Prematurely collapsing ρ_x to a single δ_{m_j} is an instance of Projection Supremacy on the interpretation layer: the system treats one admissibility class as complete, discarding the remaining $\sum_{i \neq j} w_i$ of the admissibility structure. By Theorem* **??**, *this produces unmodulated errors when the true state does not lie in class $[x_j]$.*

Remark 16.17. Anderson's claim that "brittleness in AI is the dual of overconfidence in human Intuition" is now a corollary of Theorem **??** and Proposition **??**. Logic-based systems and overconfident humans both prematurely collapse interpretation bundles. The result is identical: unmodulated errors at the edges of competence.

16.9 Robustness and Reliability as Emergent Properties

Anderson identifies robustness (tolerance of erroneous input) and reliability (tolerance of internal errors) as two aspects of the same mechanism.

Proposition 16.18 (Robustness from Continuous Familiarity Potential)**.** *The smoothed familiarity potential ϕ_C^ε is Lipschitz on all of $\mathcal{X}$, so $\mathrm{CC}^\varepsilon(x)$ varies continuously with x. Small perturbations to the input — whether from noisy observations or* *internal computation errors — produce small changes in the system's confidence measure. The system degrades continuously rather than catastrophically.*

Remark 16.19. The reason robustness and reliability are "the same mechanism" in Anderson's framework is that both arise from the continuous structure of the admissibility geometry. Internal errors are perturbations to x in the computation graph; input errors are perturbations to x in the observation. In both cases the smoothed familiarity potential provides a continuous confidence signal, and the CLIO-trained projection degrades gracefully.

16.10 The Artificial Intuition Variational Principle

Theorem 16.20 (Artificial Intuition Variational Principle)**.** *The Artificial Intuition programme is the solution to:*

$$\pi^* = \arg\min_{\pi \text{ exact minimal}} C(\mathcal{M}) \quad \textit{subject to} \quad \mathbb{E}_x\big[d_H\big(\mathcal{A}(x), \mathcal{A}_{\mathcal{M}}(\pi(x))\big)\big] < \varepsilon.$$

By Theorem **??**, *the unique solution is $\pi_{\mathcal{A}} : \mathcal{X} \to \mathcal{M}_0$. This variational principle subsumes all of Anderson's named properties of Artificial Intuition: theory-free operation (no symbolic model in the problem), prediction (minimizing L_P is equivalent to minimizing admissibility error), semantics (the solution $\mathcal{M}_0$ is the semantic space of Proposition* **??**), *ambiguity tolerance (the interpretation bundle is well-defined on $\mathcal{M}_0$), novelty (departure from $\mathcal{C} \subset \mathcal{X}$ measured by $\mathrm{Nov}^{\varepsilon}$), robustness (continuity of $\phi^{\varepsilon}_{\mathcal{C}}$ gives graceful degradation), and corpus-driven improvement ($\pi_n \to \pi_{\mathcal{A}}$ as $|\mathcal{C}_n| \to \infty$).*

Proof. The variational statement is Theorem **??** restated as an optimization. Each subsumption claim follows from the relevant proposition or theorem cited above. □

Remark 16.21. The most important sentence in Anderson's 2007 writing may not be "Intelligence is Prediction" but "Intuition operates on events, not theories," because that sentence is closest to the content of the Priority Theorem: the trajectory space and admissibility operator exist and determine the quotient before any reasoning on that quotient is possible. Anderson was describing the ordering $\mathcal{X} \to \mathcal{M}_0 \to R : \mathcal{M}_0 \to \mathcal{M}_0$ in the only language available to her at the time. The variational principle above is the mathematical expression of the same ordering.

CHAPTER 17

THE CALVIN–ANDERSON–FEYERABEND LINEAGE

> Science was created to stop people from overrating correlations and jumping to erroneous conclusions on scant evidence...But now we suddenly have Machine Learning that performs cognitive tasks at useful levels using exactly a Holistic Stance.
>
> *Monica Anderson,* The Red Pill of Machine Learning

17.1 Overview

The admissibility quotient construction did not arrive without intellectual predecessors. Three thinkers identified different aspects of the same underlying geometric constraint: William Calvin from cortical biology, Monica Anderson from AI engineering, and Paul Feyerabend from the philosophy of science. None of them had the mathematical framework. Each identified one face of a structure that the admissibility quotient makes precise.

17.2 William Calvin: The Uphill Direction

17.2.1 The Biological Problem

William Calvin's investigation concerns the relationship between the physical substrate of cognition and the phenomenology of thought. The brain is a physical system governed by electrochemical gradients, synaptic weights, and local connectivity: low-level descriptions that concern the implementation of cognition, not its content.

The puzzle Calvin poses is directionality. Physical causation flows from current states to future states, from simpler descriptions to their consequences. Yet cognition extracts higher-order structure from lower-order sensory streams: from pixels to edges to objects to scenes to meanings. This is movement in the opposite direction — a river flowing uphill.

The metaphor is precise. A river flows downhill because gravity selects the direction of least resistance in physical space. The uphill flow of cognition is driven by something else. Calvin's answer, drawing on Gerald Edelman's Neural Darwinism, is selection: repeated competitive reinforcement of neural groups that consistently produce useful reductions of their input.

17.2.2 The Mathematical Realization

In the present framework, Calvin's uphill river is the construction of the admissibility quotient hierarchy. At each level, the admissibility quotient collapses states that are indistinguishable from the perspective of accessible futures at that level, producing a coarser space with less information and a higher-level representational vocabulary.

Proposition 17.1 (Calvin's River as Quotient Hierarchy)**.** *The recursive abstraction sequence* $\mathcal{X}_0 \to \mathcal{X}_1 \to \mathcal{X}_2 \to \cdots$ *of Theorem* **??** *is the formal realization of Calvin's uphill river. Each quotient map moves from a richer, more particular space to a coarser, more abstract space. The river flows uphill because each step preserves exactly the distinctions relevant to accessible futures and discards the rest.*

The selectionist mechanism Calvin proposes is the biological correlate of the CLIO operator: neural groups making admissibility-relevant distinctions are reinforced; those that do not are suppressed. Over time the population of active representations converges toward the admissibility quotient of the organism's environment.

Proposition 17.2 (Stabilization of the Quotient Hierarchy)**.** *If* $\mathcal{X}$ *is compact and* $\mathcal{A}$ *is Hausdorff-continuous, the sequence* $(I(\mathcal{X}_n))_{n\geq 0}$ *is non-negative and non-increasing, hence convergent. The projective limit* $\mathcal{X}_\infty = \varprojlim \mathcal{X}_n$ *exists as a compact Hausdorff space, and every remaining distinction in* $\mathcal{X}_\infty$ *is admissibility-relevant.*

Proof. Non-increase: Theorem **??**. Non-negativity: mutual information is non-negative. Convergence: monotone convergence theorem. Existence of projective limit: projective limits of compact Hausdorff spaces along continuous surjections exist and are compact Hausdorff (Tychonoff and the universal property of projective limits). □

17.3 Monica Anderson: The Mechanism and the Epistemological Turn

17.3.1 The AI Engineering Context

Anderson developed her framework while trying to build systems that understand natural language. Every available method required a human to first understand the language and encode that understanding as a program — creating the circularity corollary: the thing the system was supposed to produce was required as input to its construction.

Her solution was to abandon pre-specified Models entirely. If corpus exposure produces Understanding in humans, there is no principled reason it cannot do so in machines, provided the learning mechanism is appropriate. This is the engineering origin of the Holism definition: not a metaphysical commitment but a practical consequence of the circularity diagnosis.

17.3.2 Epistemic Reduction as Pre-Scientific

Anderson's central philosophical contribution is recognizing that Epistemic Reduction operates at a level more fundamental than science. Science uses Reduction but cannot theorize it from within, because the operation that produces models is not itself a model.

In the present framework, Anderson's pre-scientific Reduction is the admissibility quotient construction. It is pre-scientific in exactly her sense: derivable from the admissibility operator alone, without reference to scientific methodology, formal languages, or a priori domain knowledge. The admissibility quotient exists independently of whether any scientist has studied the domain.

17.3.3 The 2012 Inflection as Experimental Confirmation

Anderson identified the 2012 deep learning inflection as experimental evidence for her framework. Before 2012, dominant AI required human-specified Models. After 2012, systems learning from corpus exposure outperformed model-based systems across multiple domains.

Theorem **??** makes this experimental observation into a theorem: given sufficient data and a CLIO-consistent objective, a learned system converges in probability to the admissibility quotient. The 2012 inflection is the moment when corpus sizes and architectures became adequate for this convergence to produce practically useful results. Anderson was not merely right about what happened. She was right about why, and the present framework makes that "why" precise.

17.4 Paul Feyerabend: The Warning Against Premature Closure

17.4.1 The Problem with Method

Feyerabend's *Against Method* is commonly read as arguing that science has no rational method. His actual target is more specific: no fixed method can remain adequate as science develops, because scientific discovery frequently requires recognizing phenomena that fall outside the current method's field of view. A method powerful today may become a prison tomorrow. Its power causes institutionalization as the only permissible approach, and problems it cannot solve become invisible.

Feyerabend's examples — Galileo against Aristotle, wave mechanics, continental drift — are cases where progress required violating a methodological rule or retaining a hypothesis against apparently decisive refutation. In each case the rule was a projection: a fixed decision about which features were relevant, institutionalized to the point where its implicit assumptions were invisible.

17.4.2 The Impossibility Result

Theorem **??** is the precise statement of Feyerabend's insight: a fixed projection cannot make discarded features salient through internal operations on the codomain. This is not a sociological claim but an information-theoretic impossibility. The information is absent from $\mathcal{M}$, and no endomorphism of $\mathcal{M}$ can reconstruct it.

The Aristotelian framework was a projection that discarded features of motion that Galilean mechanics required. Within the Aristotelian projection, those features were not merely unnoticed; they were not representable. No operation internal to Aristotelian mechanics could make them visible. Only a change of projection — a new compression — could reveal them.

17.4.3 Anderson and Feyerabend as Complements

Anderson asks: how should cognitive systems be built so they can discover their own salience structure? Answer: defer the projection until the corpus determines it. Feyerabend asks: why do methodological constraints sometimes prevent scientific progress? Answer: because they are projections, and premature projections prevent the discovery of features visible only at levels the current projection cannot reach.

Both arguments are instances of the Admissible Discovery Principle (Definition **??**). RSVP provides the unifying geometry: premature projection reduces the admissible exploration volume, preventing discovery of features whose salience is only revealed at admissibility levels that the premature projection cannot reach.

17.5 Three Approaches to the Same Constraint

Theorem 17.3 (Recursive Abstraction Theorem). *In the metric setting, define a sequence of trajectory spaces by $\mathcal{X}_0 = \mathcal{X}$ and*

$$\mathcal{X}_{n+1} = \mathcal{X}_n / {\sim_{\mathcal{A}n}},$$

where $x \sim_{\mathcal{A}n} y$ iff $\mathcal{A}_n(x) = \mathcal{A}_n(y)$ and $\mathcal{A}_n$ is the admissibility operator at level n, with $\mathcal{A}_{n+1}$ induced by the quotient. Then:

1. *Each $\mathcal{X}_{n+1}$ is a quotient of $\mathcal{X}_n$: there is a canonical continuous surjection $\pi_{\mathcal{A}}^{(n)} : \mathcal{X}_n \to \mathcal{X}_{n+1}$.*

2. *The mutual information is monotonically non-increasing:*

$$I(\mathcal{X}_{n+1}) \;\leq\; I(\mathcal{X}_n).$$

3. *The sequence $(\mathcal{X}_n)_{n\geq 0}$ forms a projective system of compact Hausdorff spaces.*

Proof. For (1): each $\mathcal{X}_{n+1}$ is the admissibility quotient of $\mathcal{X}_n$ by Theorem **??** applied at level n.

For (2): $\pi_{\mathcal{A}}^{(n)} : \mathcal{X}_n \to \mathcal{X}_{n+1}$ is a surjective quotient map. The data processing inequality for quotient maps states that mutual information cannot increase under a surjection, since $\mathcal{X}_{n+1}$ is a function of $\mathcal{X}_n$. Therefore $I(\mathcal{X}_{n+1}) \leq I(\mathcal{X}_n)$.

For (3): the compatible family of surjections $\pi_{\mathcal{A}}^{(n)}$ satisfies the compatibility condition $\pi_{\mathcal{A}}^{(m)} = \pi_{\mathcal{A}}^{(n)} \circ \cdots \circ \pi_{\mathcal{A}}^{(m-1)}$ for $m > n$, which is the projective system condition. □

Remark 17.4. Theorem **??** gives a formal meaning to Calvin's metaphor. Physical causation moves forward in state space; abstraction moves upward through a hierarchy of quotient spaces with decreasing mutual information. At each level, the admissibility quotient discards the distinctions that do not matter for accessible futures at that level. The river flows uphill because each quotient stage produces a strictly coarser space that retains strictly less information while preserving the admissibility structure relevant at that level.

17.6 Anderson: The Mechanism

Anderson's contribution is to name and characterize the mechanism that performs the uphill construction. Epistemic Reduction is the exact minimal admissibility-preserving projection from $\mathcal{X}$ to $\mathcal{M}$. The Anderson Projection Theorem establishes that this is the admissibility quotient, and the Recursive Abstraction Theorem gives it depth structure: deep learning is deep because Reduction must be performed at multiple levels of the hierarchy, each level discarding what it can recognize as irrelevant, and passing the rest upward.

17.7 Feyerabend: The Warning Against Premature Projection

Paul Feyerabend's argument in *Against Method* is that no fixed methodology can account for actual scientific discovery, because methodological rules are projections and premature projections exclude phenomena that have not yet been identified as salient.

Theorem 17.5 (Feyerabend–Anderson Constraint Theorem). *Let $\pi : \mathcal{X} \to \mathcal{M}$ be a fixed projection. Any feature f not encoded in $\mathcal{M}$ (i.e., discarded by π) cannot become salient through operations $R : \mathcal{M} \to \mathcal{M}$ internal to $\mathcal{M}$.*

Proof. Operations internal to $\mathcal{M}$ are endomorphisms of $\mathcal{M}$. If f has been discarded by π, then f carries no information in $\mathcal{M}$. No sequence $R_1 \circ \cdots \circ R_n : \mathcal{M} \to \mathcal{M}$ can reconstruct information absent from $\mathcal{M}$, since each R_i maps $\mathcal{M}$ to $\mathcal{M}$. □

Remark 17.6. Theorem **??** is essentially an information-theoretic impossibility result: information discarded by a projection cannot be reconstructed by endomorphisms on the codomain. It is arguably the most defensible formalization of Feyerabend's complaint in the document, and it is independent of RSVP. The problem is not method as such but Projection Supremacy applied too early: fixing a projection before the corpus is sufficient to determine which features are admissibility-sensitive.

17.8 The Admissible Discovery Principle

Definition 17.7 (Admissible Discovery Principle)**.** A system satisfies the Admissible Discovery Principle if it defers commitment to a representational manifold until the corpus is sufficient to determine which features are admissibility-sensitive, rather than fixing a projection in advance of the data.

Calvin's uphill river is the search for the admissibility quotient at each level of the recursive hierarchy. Anderson's Epistemic Reduction is its construction at each level. Feyerabend's anarchism is the warning not to freeze any level prematurely. RSVP provides the geometric optimization problem that unifies all three.

Chapter 18

TARTAN, the Wisdom Salon, and Collective Reduction

18.1 The Wisdom Salon as Distributed Epistemic Reduction

Anderson's Wisdom Salon protocol implements distributed Epistemic Reduction. Each table performs a local reduction of a complex question to a compressed semantic residue (a Grain of Wisdom). The residue passes to adjacent tables as boundary data. The harvest assembles local reductions into a global understanding.

18.2 Sheaf-Theoretic Formalization

Definition 18.1 (Admissibility Presheaf)**.** Let B be the social-epistemic base space covered by open sets $\{U_i\}$ (conversational contexts). Define the admissibility presheaf A by assigning to each U_i the set $A(U_i)$ of admissibility-preserving semantic reductions on U_i, with restriction maps $\rho_{ij} : A(U_i) \to A(U_i \cap U_j)$ given by restriction of reductions to the overlap.

Definition 18.2 (Conversational Tile and Grain of Wisdom)**.** A tile T_i computes a local section $s_i \in A(U_i)$: a local admissibility-preserving reduction of the conversational fragment over U_i. A Grain of Wisdom is the compressed residue $G_i = s_i(U_i)$.

Theorem 18.3 (TARTAN Descent Theorem)**.** *If:*

(D1) *Local reductions agree on overlaps:* $s_i|_{U_i \cap U_j} = s_j|_{U_i \cap U_j}$ *for all* i, j*; and*

(D2) *The admissibility presheaf* A *satisfies the sheaf axioms* (*locality and gluing*)*;*

then there exists a unique global section $s \in A(B)$ *such that* $s|_{U_i} = s_i$ *for all* i*, and*

$$\mathrm{colim}_i\, s_i \;\cong\; s.$$

Proof. Under (D2), the sheaf gluing axiom directly gives existence and uniqueness of a global section compatible with all local sections that agree on overlaps.

Condition (D1) is precisely the overlap-consistency hypothesis of the gluing axiom. The colimit identification follows from the universal property of the colimit and the uniqueness of the global section. □

Remark 18.4. The previous draft stated that the colimit of local quotients "approximates the global admissibility quotient." That claim is stronger than what was established: a colimit of local quotients need not equal a quotient of the global space in general. Theorem **??** replaces the approximation language with a precise condition: when the admissibility presheaf is a genuine sheaf and local reductions are consistent on overlaps, the colimit is isomorphic to the unique global section guaranteed by the gluing axiom. The sheaf condition (D2) is the non-trivial hypothesis; verifying it in a given application requires checking that local admissibility reductions restrict and extend consistently.

Proposition 18.5 (Good Ideas are Stable Sections)**.** *A Grain G with low holonomy (content remaining consistent after parallel transport across B) appears in the global section s. A Grain with high holonomy (conflicting when transported into different conversational contexts) fails the gluing condition (D1) and does not appear in s.*

Remark 18.6. Proposition **??** formalizes Anderson's observation about the filtering power of the World Café Protocol. The filtering power comes from transport stability, not voting. Wisdom is what survives parallel transport.

CHAPTER 19

CONNECTIONS

19.1 Anderson and Kahneman

System 1 in Kahneman's framework corresponds to the CLIO approximation of the admissibility quotient: rapid, parallel, below conscious access, improving with practice. System 2 corresponds to $R : \mathcal{M} \to \mathcal{M}$: sequential, deliberate, operating inside the manifold produced by System 1. Theorem **??** is the formal version of Kahneman's observation that System 2 depends on System 1 outputs.

19.2 Anderson and the Frame Problem

The Frame Problem asks how a cognitive system specifies what remains unchanged after an action. In admissibility geometry this is the problem of computing $\mathcal{A}(x')$ from $\mathcal{A}(x)$ after a transition $x \to x'$: which admissible futures survive a state change. A system with a correct admissibility operator solves it automatically: unchanged admissible futures appear in both sets.

19.3 Anderson and the Author's Prior Work

The author's *Beyond Prediction Error* develops admissibility-based ecological cognition from dynamical starting points. *Function Survives Collapse* develops admissibility preservation through secure computation, biological regulation, and typographic systems. Both converge on the same principle: cognition is not prediction but the maintenance of viable futures under projection. Anderson's Epistemic Reduction is the epistemological face; the admissibility quotient is its geometric face.

19.4 The Yoneda Connection

Anderson's aphorism "you are known by the company you keep" is her gloss on the Yoneda Lemma: a mathematical object is fully determined by its relationships to all other objects. In the manifold setting, the identity of $[x] \in \mathcal{M}_0$

is fully determined by $\mathcal{A}_{\mathcal{M}_0}([x]) = \pi_{\mathcal{A}}(\mathcal{A}(x))$. Corpus Congruence is the empirical approximation of this Yoneda representation via the smoothed familiarity potential.

CONCLUSION

The thesis of this monograph is that Monica Anderson's epistemological program is derivable from the geometry of admissible projection, not merely compatible with it. The programme spans two decades of Anderson's writing, from the 2007 *Artificial Intuition* site through the *Little Pills* essays, and the present monograph covers both.

The framework rests on the admissibility quotient construction. Given a compact Hausdorff trajectory space $\mathcal{X}$ and a Hausdorff-continuous admissibility operator $\mathcal{A}$, the quotient $\mathcal{M}_0 = \mathcal{X}/{\sim_{\mathcal{A}}}$ exists as a compact Hausdorff space and is the unique information-minimal admissibility-preserving representation of $\mathcal{X}$ among exact minimal projections. This is established by the Existence Theorem, the Universal Property, and the Information Minimality Theorem.

From this construction, Anderson's entire epistemological program follows as a series of theorems and corollaries. The Identification Theorem (Theorem **??**) establishes the most philosophically important result: every operation Anderson attributes to Understanding — salience detection, corpus congruence, novelty recognition, abstraction formation, graceful degradation, model-free operation, corpus-driven improvement, and priority over Reasoning — is realized by a property of the admissibility quotient and its associated fields. The identification is not imposed on Anderson's ideas from outside; it is extracted from them by asking what mathematical structure would realize each property she attributes to Understanding. The answer, in each case, is the admissibility quotient or one of its associated fields.

The Anderson Projection Theorem establishes the four-way equivalence between Epistemic Reduction, exact minimal admissibility-preserving projection, factoring through the quotient, and salience-filtering. The Priority Theorem derives Anderson's most famous claim as a structural tautology. The Brittleness Theorem explains classical AI failure as Projection Supremacy. The Deep Learning Convergence Theorem derives Anderson's claim about why deep learning works: under CLIO consistency conditions, learned network projections converge in probability to the admissibility quotient.

The Artificial Intuition chapter (Chapter **??**) addresses the 2007 framework directly and provides the chapter that was previously missing from the monograph. Anderson's claim that "Intuition operates on events, not theories" is reinterpreted as "the projection precedes reasoning," which is the Priority Theorem. Her "Intelligence is Prediction" is reinterpreted as admissibility estimation, not point forecasting. Her Bizarre Domains taxonomy is reinterpreted as four failure modes of admissibility-preserving projection, unifying what had appeared to be independent categories into a single geometric obstruction: non-decomposability of the admissibility operator. The Artificial Intuition Variational Principle (Theorem **??**) then compresses Anderson's entire 2007 programme into a single optimization statement whose unique solution is the admissibility quotient.

The Calvin–Anderson–Feyerabend synthesis provides the historical and philosophical context. Calvin's uphill river is the quotient hierarchy; its stabilization theorem gives the fixed point that cognitive systems approach. Anderson's epistemological program is the practical description of the quotient construction from an engineering perspective. Feyerabend's impossibility result is Theorem **??**: information discarded by a premature projection cannot be recovered by internal operations. The Admissible Discovery Principle unifies all three.

The CLIO Convergence theorem is conditional on three explicit hypotheses. The Local Extension Stability theorem is one-directional. The full Curvature–Admissibility Threshold is identified as a conjecture with its missing ingredients named. These are not weaknesses but honest boundaries. The core framework is exact; the conditional results are useful under stated conditions; the conjecture identifies the deepest remaining open problem.

The Semantic Emergence theorem (Theorem **??**) establishes the deepest link in the monograph: prediction itself generates the admissibility quotient hierarchy. Stable prediction equivalence classes are the semantic manifold. The full implication chain — Prediction $\Rightarrow$ Semantics $\Rightarrow$ Understanding $\Rightarrow$ Reasoning — that Anderson proposed in 2007 is now provable within the framework. The three arrows correspond respectively to the Semantic Emergence theorem, the Identification Theorem, and the Priority Theorem.

Three further results in the Brittleness chapter extend the framework to collective and dynamic phenomena. Interpretive Attractors formalize the observation that projections concentrate into self-reinforcing frameworks. The Community Semantic Manifold proposition shows that understanding is a collective geometric structure maintained through shared corpora, explaining why identical events can produce radically different conclusions in different communities. Projection Capture formalizes the dynamic failure mode that Projection Supremacy does not cover: a projection that has become insulated from revision by its own prior successes, losing the ability to discover new saliences by the Feyerabend–Anderson Constraint.

Anderson's insight — that intelligence is fundamentally about choosing what can be ignored without losing what matters — is a variational principle: find the exact minimal admissibility-preserving projection. The admissibility quotient is its canonical solution. The Identification Theorem shows that Anderson already knew this, expressed in a vocabulary adequate to her engineering context. The present monograph provides the geometry that explains why she was right.

Appendix A

Glossary

Admissibility Quotient $\mathcal{M}_0 = \mathcal{X}/{\sim_{\mathcal{A}}}$: the canonical minimal representational manifold identifying states with identical admissible futures.

Admissibility-Preserving Projection (weak) Commutes with admissible dynamics: $\pi(\mathcal{A}(x)) = \mathcal{A}_{\mathcal{M}}(\pi(x))$. May retain extra information.

Admissible Discovery Principle Defer commitment to a representational manifold until the corpus determines which features are admissibility-sensitive.

Artificial General Learner A system learning competence across wide domains through data exposure, without antecedent domain-specific Understanding encoded as programs.

Bizarre Domain A problem domain in which no Model can be simultaneously tractable and faithful.

CLIO Machine-learning operator approximating the admissibility quotient by minimizing L_{CLIO}; convergent under (C1)–(C3).

Corpus Congruence $\mathrm{CC}^{\varepsilon}(x) = e^{-d(x,\mathcal{C})/(\varepsilon + d(x,\mathcal{C}))}$.

Epistemic Reduction Salience-preserving projection from rich experience to structured representation; equivalent to exact minimal admissibility-preserving projection.

Exact Minimal Admissibility-Preserving Projection $\pi(x) = \pi(y)$ iff $\mathcal{A}(x) = \mathcal{A}(y)$: no redundant information retained and no admissibility distinction collapsed.

Familiarity Potential (smoothed) $\phi^{\varepsilon}_{\mathcal{C}}(x) = -\log(\varepsilon + d(x,\mathcal{C}))$.

Holism Avoidance of a priori Models; allowing the projection to be learned from data.

Model A context-free simplification of reality.

Projection Supremacy Applying R without detecting projection error.

Reductionism The use of Models.

Salience (metric) A feature whose deletion changes some admissible future cone.

Salience (smooth RSVP) A feature for which $\nabla_f S(x) \neq 0$; a computable proxy, not equivalent to the set-theoretic criterion.

TARTAN Tiled Admissibility-Reduction Traversal Architecture; distributed CLIO assembled via sheaf descent.

Understanding Construction of $\pi : \mathcal{X} \to \mathcal{M}$.

Reasoning Operation $R : \mathcal{M} \to \mathcal{M}$ on the manifold produced by Understanding.

APPENDIX B

NOTATION REFERENCE

Symbol	Meaning
$\mathcal{X}$, $(\mathcal{X}, d_{\mathcal{X}})$	Trajectory space (compact metric)
$\mathcal{M}$, $\mathcal{M}_0$	Representational manifold; admissibility quotient
$\mathcal{X}_n$	Trajectory space at recursive abstraction level n
$\pi_{\mathcal{A}}$	Canonical projection to $\mathcal{M}_0$
$\pi_{\mathcal{A}}^{(n)}$	Canonical projection at level n
$\pi : \mathcal{X} \to \mathcal{M}$	General projection
$\mathcal{A}(x)$	Admissible future cone of x (subset of $\mathcal{X}$)
$\mathcal{A}_{\mathcal{M}}(m)$	Induced admissibility on $\mathcal{M}$: $\pi(\mathcal{A}(x))$ for $\pi(x) = m$
$\sim_{\mathcal{A}}$	Admissibility equivalence: $\mathcal{A}(x) = \mathcal{A}(y)$
d_H	Hausdorff distance on compact subsets
$S(x)$	Accessibility entropy: $\log \mu(\mathcal{A}(x))$
$\phi_{\mathcal{C}}^{\varepsilon}$	Smoothed familiarity potential: $-\log(\varepsilon + d(x, \mathcal{C}))$
$\nabla \phi_{\mathcal{C}}^{\varepsilon}$	Inferential flow field (RSVP $\mathbf{v}$)
$\mathrm{CC}^{\varepsilon}(x)$	Corpus Congruence (smoothed)
$\mathrm{Nov}^{\varepsilon}(x)$	Novelty: $1 - \mathrm{CC}^{\varepsilon}(x)$
ε	Smoothing parameter for familiarity potential
$\mathcal{C}$	Training corpus (compact subset of $\mathcal{X}$)
L_{CLIO}	CLIO reduction functional
$R : \mathcal{M} \to \mathcal{M}$	Reasoning operator
$L_u : \mathcal{M}_0 \to \mathcal{M}_0'$	Learning extension operator
K_{max}	Curvature bound for Local Extension Stability
$A(U_i)$	Admissibility presheaf over U_i
$s_i \in A(U_i)$	Local section (Grain of Wisdom)
ρ_{ij}	Restriction map between local sections
$\mathrm{colim}_i\, s_i$	Harvest (colimit of local sections)
$I(\mathcal{X}; \mathcal{M})$	Mutual information between $\mathcal{X}$ and $\mathcal{M}$

APPENDIX C

SUMMARY OF MAIN RESULTS

Lemma 5.1 *Closure.* Hausdorff-continuity of $\mathcal{A}$ implies $\sim_{\mathcal{A}}$ is a closed equivalence relation. PROVED.

Theorem 6.1 *Existence of the Admissibility Quotient.* Under compactness and Hausdorff-continuity, $\mathcal{M}_0 = \mathcal{X}/{\sim_{\mathcal{A}}}$ is compact Hausdorff with a unique exact minimal admissibility-preserving canonical projection $\pi_{\mathcal{A}}$. PROVED.

Theorem 6.2 *Universal Property.* Every exact minimal admissibility-preserving projection factors through $\mathcal{M}_0$ via an injection; $\mathcal{M}_0$ is initial in the relevant category. PROVED.

Theorem 6.3 *Information Minimality.* All exact minimal admissibility-preserving projections carry the same mutual information $I(\mathcal{X}; \mathcal{M}_0)$. The quotient is the unique such representation up to isomorphism. PROVED.

Proposition 6.4 *Entropy Preservation as Necessary Condition.* Admissibility preservation implies entropy preservation. The converse requires additional assumptions; in general equal entropy does not imply equal admissible future sets. PROVED; CONVERSE FAILURE SHOWN BY COUNTEREXAMPLE.

Theorem 7.1 *Salience–Admissibility Equivalence.* A feature is salient iff its deletion changes some admissible future cone. PROVED (METRIC SETTING).

Proposition 7.2 *Salience as Entropy Gradient.* $\nabla_f S(x) \neq 0$ is a computable proxy for salience. Not equivalent to the set-theoretic criterion. PROVED (SMOOTH RSVP SETTING).

Proposition 7.3 *Gradient of Familiarity Potential.* $\nabla \phi^{\varepsilon}_{\mathcal{C}} = -\nabla d(\,\cdot\,, \mathcal{C}) \,/\, (\varepsilon + d(\,\cdot\,, \mathcal{C}))$ on any Riemannian manifold with differentiable distance function. PROVED.

Theorem 9.1 *Priority of Understanding.* Reasoning requires a prior Understanding; Understanding does not require Reasoning. PROVED (STRUCTURAL TAUTOLOGY).

Theorem 10.1 *The Anderson Projection Theorem.* Four characterizations of Epistemic Reduction are equivalent. Entropy preservation is a consequence, not an equivalent. PROVED (METRIC SETTING; ENTROPY COROLLARY IN SMOOTH RSVP).

Theorem 11.1 *CLIO Consistency.* Empirical CLIO minimizers converge in probability to $\pi_{\mathcal{A}}$ under conditions (C1)–(C3). PROVED CONDITIONALLY.

Theorem 12.1 *Local Extension Stability.* Bounded curvature in S near m_u implies locally stable deformation under learning extension L_u. Converse not proved. PROVED (ONE DIRECTION ONLY).

Conjecture 12.2 *Curvature–Admissibility Threshold.* Admissibility-compatible learning holds iff Hessian of S is bounded by a threshold determined by global geometry of $\mathcal{M}_0$. Missing ingredients identified in Chapter **??**. OPEN.

Theorem 13.1 *Brittleness from Projection Supremacy.* Projection Supremacy produces unmodulated errors outside the zero-error region. PROVED.

Theorem 15.1 *Recursive Abstraction.* The sequence $(\mathcal{X}_n)$ of iterative admissibility quotients forms a projective system with monotonically decreasing mutual information. PROVED.

Theorem 15.2 *Feyerabend–Anderson Constraint.* Features discarded by a fixed projection cannot become salient through endomorphisms on the codomain. PROVED (INFORMATION-THEORETIC IMPOSSIBILITY).

Theorem 16.1 *TARTAN Descent.* Under overlap consistency (D1) and sheaf axioms (D2), the colimit of local sections is isomorphic to the unique global section. PROVED (CONSEQUENCE OF SHEAF GLUING AXIOM).

Proposition 16.2 *Stable Sections as Good Ideas.* Low-holonomy sections survive sheaf transport; high-holonomy sections fail the gluing condition. PROVED.

Theorem (Identification) *The Identification Theorem.* Every operation Anderson attributes to Understanding — salience detection, corpus congruence, novelty recognition, abstraction formation, graceful degradation, model-free operation, corpus-driven improvement, and priority over Reasoning — is realized by a property of the admissibility quotient or its associated fields. PROVED (AS COMPILATION OF PRIOR RESULTS).

Proposition (Models–Projection) *Models are Projections.* A Model in Anderson's sense corresponds to a projection $\pi : \mathcal{X} \to \mathcal{M}$. Model creation is projection specification. Discarded context is fiber information. Model-Based AI Circularity follows from Theorem **??**. PROVED.

Theorem (Depth Necessary) *Depth is Necessary for Multi-Level Salience.* When the admissibility quotient hierarchy has depth $N > 1$, no single-layer projection can replicate the filtering performed by the full hierarchy. PROVED.

Theorem (Deep Learning Convergence) *Deep Learning as Approximate Admissibility Reduction.* Under CLIO conditions (C1)–(C3), trained deep network projections converge in probability to the admissibility quotient. PROVED CONDITIONALLY (INHERITS CLIO CONDITIONS).

Proposition (Calvin's River) *Calvin's River as Quotient Hierarchy.* The recursive abstraction sequence is the formal realization of Calvin's uphill river. PROVED AS INTERPRETATION OF THEOREM **??**.

Proposition (Stabilization) *Stabilization of the Quotient Hierarchy.* The projective limit $\mathcal{X}_\infty = \varprojlim \mathcal{X}_n$ exists as a compact Hausdorff space in which every remaining distinction is admissibility-relevant. PROVED.

Definition (Artificial Intuition) Approximation of $\pi_\mathcal{A} : \mathcal{X} \to \mathcal{M}_0$ from corpus alone, with no symbolic domain theory as input. DEFINITION.

Theorem (Theory-Free Learning) Under CLIO conditions (C1)–(C3), corpus-driven empirical projections converge to $\pi_\mathcal{A}$ without specifying a symbolic theory. Formalizes "Intuition learns directly from events." PROVED CONDITIONALLY (INHERITS CLIO CONDITIONS).

Proposition (Emergent Semantics) Semantic categories are admissibility attractors in the quotient hierarchy: stable regions where accessible future structure is approximately constant. PROVED.

Proposition (Bizarre Domains as Projection Failure Modes) Anderson's four Bizarre Domain categories (Chaos, Holism, Ambiguity, Emergence) correspond to four distinct failure modes of admissibility-preserving projection, all arising from non-decomposability of $\mathcal{A}$. PROVED.

Theorem (Artificial Intuition Variational Principle) The AI programme is the variational problem $\arg\min C(\mathcal{M})$ subject to admissibility error $< \varepsilon$. The unique solution is $\pi_\mathcal{A}$, and the principle subsumes all of Anderson's named AI properties. PROVED.

Proposition (Premature Collapse as Projection Supremacy) Prematurely selecting a single interpretation from an interpretation bundle is an instance of Projection Supremacy; produces unmodulated errors by Theorem **??**. PROVED.

Theorem (Semantic Emergence by Prediction Quotients) Under CLIO conditions, the prediction-equivalence relation $\sim_P$ converges in probability to the admissibility equivalence $\sim_\mathcal{A}$. Consequently: prediction quotient $\to \mathcal{M}_0$; stable semantic categories are attractors of prediction dynamics; Semantics = Stable Equivalence Classes of Prediction. Establishes the full chain Prediction $\Rightarrow$ Semantics $\Rightarrow$ Understanding $\Rightarrow$ Reasoning. PROVED CONDITIONALLY (INHERITS CLIO CONDITIONS).

Definition (Interpretive Attractor) A compact region of $\mathcal{M}_0$ with high internal accessibility and high boundary crossing cost; formalizes stable interpretive frameworks. DEFINITION.

Proposition (Frameworks as Self-Reinforcing Projections) A projection concentrated in an interpretive attractor assimilates new inputs into the existing framework rather than expanding the projection. PROVED.

Proposition (Community Semantic Manifold) Under repeated exchange, agents converge toward a shared admissibility quotient of the community corpus. Identical events projected through different community manifolds occupy different salience locations. PROVED.

Definition (Projection Capture) The dynamic analogue of Projection Supremacy: the CLIO update gradient is suppressed by prior reinforcement, insulating the projection from revision even under positive projection error. DEFINITION.

Proposition (Projection Capture as Dynamic Projection Supremacy) Projection capture is Projection Supremacy persisted dynamically; the system loses the ability to update π and therefore cannot discover new saliences, by Theorem **??**. PROVED.

BIBLIOGRAPHY

Anderson 2007 Anderson, Monica. *Artificial Intuition: A New Possible Path to Artificial Intelligence.* `https://artificial-intuition.com`, 2007. [Archived primary source; original site covering Bizarre Domains, Intuition vs. Logic, prediction as intelligence, and the Tradeoff framework. Cited throughout as the 2007 text.]

Anderson 2017a Anderson, Monica. *Why AI Works: Intelligence = Understanding + Reasoning.* Artificial Understanding (Medium), 2017.

Anderson 2017b Anderson, Monica. *Our Greatest Invention: Model-Based Problem Solving.* Artificial Understanding (Medium), 2017.

Anderson 2017c Anderson, Monica. *Two Dirty Words: Reductionism and Holism.* Artificial Understanding (Medium), 2017.

Anderson 2017d Anderson, Monica. *Reduction: An Introduction.* Artificial Understanding (Medium), 2017.

Anderson 2018 Anderson, Monica. *Why Deep Learning Works: Deep Learning Performs Epistemic Reduction.* Experimental Epistemology, 2018. `https://experimental-epistemology.ai`

Anderson 2022a Anderson, Monica. *The Red Pill of Machine Learning.* Experimental Epistemology, 2022.

Anderson 2022b Anderson, Monica. *Corpus Congruence.* Experimental Epistemology, 2022.

Anderson 2022c Anderson, Monica. *Model Free AI: Don't Model the World—Just Model the Mind.* Experimental Epistemology, 2022.

Anderson 2022d Anderson, Monica. *Experimental Epistemology for AI.* Experimental Epistemology, 2022.

Anderson 2022e Anderson, Monica. *The Wisdom Salon.* Experimental Epistemology, 2022.

Anderson/Flyxion 2022 Anderson, Monica; edited by Flyxion. *Monica's Little Pills.* Standard Galactic Edition, 2022.

Calvin 1996 Calvin, William H. *How Brains Think: Evolving Intelligence, Then and Now.* Basic Books, New York, 1996.

Edelman 1987 Edelman, Gerald M. *Neural Darwinism: The Theory of Neuronal Group Selection.* Basic Books, New York, 1987.

Feyerabend 1975 Feyerabend, Paul. *Against Method: Outline of an Anarchistic Theory of Knowledge.* New Left Books, London, 1975.

Flyxion 2026a Flyxion. *Adjunction and Irreversibility: RSVP Field Theory.* Independent monograph, 2026. `https://standardgalactic.github.io/calculus/computation/adjunction_and_irreversibility.pdf`

Flyxion 2026b Flyxion. *The Projection Crisis: Constraint, Trajectory, and Irreversibility.* Independent monograph, 2026. `https://standardgalactic.github.io/playfloor/epistemology/projection_crisis.pdf`

Flyxion 2026c Flyxion. *Beyond Prediction Error: Admissibility-Based Ecological Cognition.* Independent monograph, 2026. `https://standardgalactic.github.io/alphabet/framework/beyond_prediction_error.pdf`

Flyxion 2026d Flyxion. *Function Survives Collapse.* Independent monograph, 2026. `https://standardgalactic.github.io/alphabet/working/function_survives_collapse.pdf`

Kahneman 2011 Kahneman, Daniel. *Thinking, Fast and Slow.* Farrar, Straus and Giroux, New York, 2011.

Krizhevsky et al. 2012 Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25, 2012.

Lee 2003 Lee, John M. *Introduction to Smooth Manifolds.* Graduate Texts in Mathematics 218. Springer, New York, 2003.

Mac Lane 1971 Mac Lane, Saunders. *Categories for the Working Mathematician.* Graduate Texts in Mathematics 5. Springer, New York, 1971.

McCarthy & Hayes 1969 McCarthy, John; Hayes, Patrick J. Some Philosophical Problems from the Standpoint of Artificial Intelligence. *Machine Intelligence*, 4:463–502, 1969.

Pirsig 1974 Pirsig, Robert M. *Zen and the Art of Motorcycle Maintenance.* William Morrow, New York, 1974.

Schrödinger 1944 Schrödinger, Erwin. *What Is Life?* Cambridge University Press, Cambridge, 1944.

Smuts 1926 Smuts, Jan Christiaan. *Holism and Evolution.* Macmillan, London, 1926.

Vapnik 1998 Vapnik, Vladimir N. *Statistical Learning Theory.* Wiley, New York, 1998.

Winston 1975 Winston, Patrick H. *The Psychology of Computer Vision.* McGraw-Hill, New York, 1975.