

# Structural Preconditions of Describable Operations:

A Methodological Framework for Analyzing Initial Conditions

Boris Kriger<sup>1</sup>

<sup>1</sup>Institute of Integrative and Interdisciplinary Research,  
`boriskrigger@interdisciplinary-institute.org`

## Abstract

This paper establishes a conditional theorem: for any process that can be described as undergoing state transitions, initial states of zero structure cannot function as operational starting points for describable processes. The argument is epistemic and operational, not metaphysical. We make no claims about metaphysical nothingness or what may exist beyond the reach of description; we claim only that the requirements of describability entail non-empty structural preconditions.

We develop two principles. The Principle of Non-Zero Initialization states: if a process is describable as having transitions, then its description presupposes a non-empty domain of differentiable states. The Principle of Medium Constraint states: any formal language or modeling framework used to represent a system constrains what can be expressed prior to any specific act of representation. Together, these yield Double Preconditioning: for systems that are both describable and representable, initial states lie at the intersection of two non-empty constraint spaces.

The paper maintains strict separation among three levels: (1) conceptual necessities following from the requirements of description, (2) empirical illustrations showing the framework's relevance to cognitive science, artificial intelligence, and mathematics, and (3) pragmatic implications. Empirical cases serve as illustrations of applicability, not as premises or evidence for the core argument.

The thesis is domain-restricted: it concerns systems amenable to description, modeling, and analysis. Questions about metaphysical origins, creation *ex nihilo*, or domains beyond description lie outside this paper's scope.

**Keywords:** structural preconditions, describability, operational constraints, initial conditions, path dependence, inductive bias, formal systems

## 1 Introduction

A well-known piece of practical wisdom states: “Grant me the serenity to accept the things I cannot change, the courage to change the things I can, and the wisdom to know the difference.”

In everyday life this distinction appears self-evident. Yet in scientific reasoning, technological design, and social planning it is repeatedly neglected. Across many domains we encounter implicit assumptions that systems can be treated as if they began from neutral, unconstrained states; that models can be constructed without inherited biases; or that institutions and algorithms can be redesigned “from scratch.” These assumptions constitute modern forms of the *tabula rasa fallacy*—the belief that beginnings can be structurally unconditioned.

The present paper proposes that the familiar wisdom about “knowing the difference” requires a formal methodological foundation. We aim to analyze, rather than merely assert, why such a distinction is necessary, how it can be rigorously articulated, and what kinds of errors follow

when it is ignored. Our goal is not to offer moral guidance but to provide an epistemic framework capable of separating predetermined constraints from modifiable degrees of freedom in any describable system.

To this end, we develop two complementary principles: the Principle of Non-Zero Initialization, which establishes that operational systems require pre-existing structure, and the Principle of Medium Constraint, which shows that any act of description imposes additional limitations. Their conjunction yields what we call *Double Preconditioning*—a formal explanation of why no process can begin from absolute neutrality within any domain open to analysis.

In this sense, the paper can be read as a methodological elaboration of an old intuition. Where practical wisdom advises us to distinguish the unchangeable from the changeable, we seek to explain why such a distinction is unavoidable, how it can be systematically drawn, and how to avoid the recurrent errors that arise when it is overlooked.

## 1.1 The Question and Its Scope

The expression *tabula rasa*—Latin for “blank slate”—has held philosophical appeal for millennia. The concept suggests that origins can be neutral, that beginning points are structurally uncommitted, and that what follows is purely a function of subsequent inputs rather than antecedent conditions.

This paper advances a conditional claim: for any system that is describable, operational, or evolvable, zero-structure states cannot function as initial conditions for describable processes. This is not a claim about metaphysical nothingness or the ultimate nature of reality. It is a claim about the requirements of describability and operation.

### SCOPE AND METHODOLOGICAL CLARIFICATION

This paper’s argument is epistemic and operational, not metaphysical.

**We claim:** The description of a process as undergoing change requires differentiable states, which constitutes non-zero structure.

**We do NOT claim:**

- That metaphysical nothingness is impossible in some absolute sense
- That zero-structure states cannot be conceived or formally represented
- Anything about domains beyond description

The argument concerns conditions for description and operation, not the ultimate nature of reality.

## 1.2 Operational vs. Logical Impossibility

We distinguish two types of impossibility:

**Logical impossibility:** A state containing internal contradiction (like a married bachelor). Such states cannot be coherently conceived.

**Operational impossibility:** A state incompatible with serving as an initial condition for any describable process. Such states may be formally representable but cannot function as starting points for description or operation.

We claim the second, not the first. One can write “ $\emptyset$ ” and work with it in formal systems. Our claim is that zero-structure states cannot function as initial conditions for describable processes—not that they cannot be represented or conceived.

## 1.3 The Logical Structure of the Argument

The core argument has the following structure:

## LOGICAL DERIVATION

1. The description of change  $\rightarrow$  requires distinguishing states (before/after)
2. Distinguishing states  $\rightarrow$  requires differentiable elements
3. Differentiable elements  $\rightarrow$  constitutes non-empty domain
4. Therefore: The description of change  $\rightarrow$  non-empty domain

This is a conditional derivation from the requirements of description, not a metaphysical claim about necessity in all possible worlds.

### 1.4 Three Levels of Analysis

This paper maintains strict separation among three levels:

**Level 1 – Conceptual necessities:** Claims following from the requirements of descriptibility. These are the core argument: if something is describable as undergoing change, then certain preconditions must obtain.

**Level 2 – Empirical illustrations:** Cases from cognitive science, AI, and mathematics illustrating that the framework captures something relevant. These are not proofs or evidence for the conceptual claims but demonstrations of applicability.

**Level 3 – Pragmatic implications:** Practical consequences for thinking about learning, AI, and modeling. These follow from applying the framework to specific contexts.

### 1.5 Paper Structure

Section 2 reviews relevant literature. Section 3 develops the formal framework with precise definitions. Section 4 presents Double Preconditioning. Section 5 examines the self-initialization objection. Section 6 addresses mathematical emptiness. Section 7 illustrates applications to AI. Section 8 catalogues objections. Section 9 discusses practical applications.

### 1.6 Methodological Character, Novel Contribution, and Cognitive Safeguard

This paper is fundamentally methodological in character. Its aim is to identify and formalize a recurrent epistemic error in the conceptualization of initial conditions across diverse fields, and to provide a principled framework for avoiding that error in future analyses. The error at issue is the widespread tendency to treat describable processes as though they could meaningfully begin from states of absolute structural neutrality. We refer to this tendency as the *tabula rasa fallacy*. The fallacy does not arise from metaphysical speculation about nothingness, but from an insufficient appreciation of the constraints that necessarily accompany any act of description or representation. By articulating these constraints through the Double Preconditioning Theorem, the paper offers a diagnostic instrument and a corrective mechanism, transforming an often implicit oversight into an explicit methodological principle.

Our approach is neither metaphysical nor broadly philosophical in a speculative sense. It advances no claims about the ultimate nature of reality, about the possibility or impossibility of absolute nothingness, or about origins in any ontological register. Instead, it is strictly epistemic and methodological. The argument concerns only the conditions under which descriptions of operational processes can be made coherent and analyzable. The central question is not what may exist in principle, but what can be meaningfully specified within a framework of representation.

The novelty of the contribution lies in three interconnected dimensions.

**First, the paper diagnoses a recurring cognitive bias that appears across disciplines.** This bias, analogous to confirmation bias in its selective disregard for inconvenient

preconditions, manifests whenever analysts implicitly assume that systems, models, or institutions can be meaningfully conceived as starting from a neutral or structureless state. In cognitive science it appears in overly simplified empiricist models that posit learning without inductive priors; in artificial intelligence it appears in claims of “unbiased” algorithms that ignore architectural and data inheritances; in the social sciences it appears in narratives of radical institutional resets that neglect path dependence and embedded constraints. These examples function only as illustrations, not as evidence for the theorem. Their role is to show how the same conceptual misframing recurs whenever the prerequisites of description are overlooked.

**Second, the framework elevates this diagnosis into a formal safeguard.** The principles advanced here are not tautological restatements of definitions. They are derived from minimal epistemic assumptions about what is required for any process to be described at all. We begin with a deliberately structure-agnostic notion of a system—anything that can be represented as undergoing transitions—and show that the act of describing such transitions necessarily entails differentiable states and a non-empty domain. The premises concern only the requirements of representation; the conclusion follows as a consequence, not as an assumption. The value of the theorem is therefore prophylactic: it provides analysts with a clear checklist for detecting when the tabula rasa fallacy has been tacitly introduced into a model or argument.

**Third, the originality of the contribution lies in its interdisciplinary synthesis and its prescriptive utility.** Elements of the argument resonate with established ideas such as inductive bias in machine learning, the poverty-of-the-stimulus argument in linguistics, and path dependence in economics. What has been missing, however, is a unifying framework that connects these insights under a single conditional theorem, while carefully separating epistemic necessities from empirical illustrations and pragmatic consequences. The tri-level structure adopted in this paper—necessities, illustrations, and implications—ensures methodological clarity and prevents the conflation of logical requirements with domain-specific facts.

The intent of this work is therefore modest but precise. By “fallacy” we mean not a logical contradiction, but a persistent methodological misframing that ignores the prerequisites of description. The theorem does not claim that absolute structural neutrality is metaphysically impossible. It claims only that such neutrality cannot function as a meaningful starting point within any describable or operational domain. In this sense, the framework serves as a form of cognitive safeguard: it disciplines inquiry by requiring explicit recognition of the constraints that necessarily precede any coherent representation.

In sum, the contribution is not a rediscovery of the obvious, but a targeted intervention against a pervasive conceptual shortcut. The conditional form of the argument preserves appropriate humility while providing actionable rigor. It strengthens methodological precision without making claims that exceed what the logic of description can support.

## 1.7 Types of Methodological Errors in Reasoning About Initial Conditions

The preceding sections have established that no describable process can be coherently modeled as beginning from a state of absolute structural neutrality. Yet this conclusion does not imply that all aspects of an initial condition are equally predetermined. In practical analysis, initial states typically contain both predetermined components and components that are genuinely open to variation. Methodological confusion arises when these domains are conflated. (For extensive illustrations of these errors across history, science, policy, psychology, and other domains, see Appendix A.)

To clarify this point, we introduce a simple notational scheme. Let any initial condition  $I$  be understood as composed of two subsets:

- $P$  – elements that are *predetermined* within the relevant descriptive framework (architectural constraints, physical laws, representational limits, inherited parameters);

- $F$  – elements that remain *formally free* or underdetermined relative to that framework (choices, adjustable parameters, contingent inputs).

Thus any describable starting point can be schematized as:

$$I = P + F \tag{1}$$

The tabula rasa fallacy consists in treating  $I$  as if  $P = \emptyset$ . The opposite error consists in treating  $F = \emptyset$ . Both distortions generate characteristic methodological mistakes.

### **Error Type I: Misattribution of Freedom (False Neutrality)**

This error occurs when elements that belong to the predetermined domain  $P$  are incorrectly treated as if they were freely selectable elements of  $F$ .

Formally:  $P \rightarrow F$

Examples include:

- Claims that an algorithm can be “unbiased” despite fixed architectural inductive biases;
- Educational theories that assume learners arrive without cognitive priors;
- Institutional proposals that promise complete resets while ignoring legal, infrastructural, or historical constraints.

The structure of the error is to imagine optionality where none exists. This is the most direct manifestation of the tabula rasa fallacy: the assumption that what is structurally fixed can be treated as if it were neutral.

### **Error Type II: Misattribution of Determination (False Fixity)**

The symmetrical mistake arises when genuinely open or adjustable components  $F$  are mistakenly classified as predetermined  $P$ :

Formally:  $F \rightarrow P$

Here the analyst denies the presence of partial tabula rasa where it in fact exists. Typical instances include:

- Arguments that social or cognitive outcomes are entirely fixed by genetics or architecture;
- Technological determinism that dismisses the possibility of alternative designs or policies;
- Modeling practices that treat tunable parameters as immutable features.

This error leads to fatalism and rigidity: the refusal to recognize degrees of freedom that are in fact available for modification.

### **Error Type III: Domain Conflation (Category Misplacement)**

The most serious methodological mistake occurs when the boundaries between  $P$  and  $F$  are not merely misclassified, but assigned to the wrong descriptive domain altogether.

Formally, if domains  $D_1$  and  $D_2$  require different partitions of  $P$  and  $F$ , the error is:

$$(P + F)_{D_1} \text{ applied to } D_2 \tag{2}$$

Examples include:

- Treating sociocultural constraints as if they were biological necessities;
- Treating computational limitations as if they were logical impossibilities;
- Assuming that constraints of a particular modeling environment represent constraints of the target system itself.

This form of error is especially damaging because it does not merely misallocate freedom and constraint; it misidentifies the very level at which the analysis is taking place. It confuses representational preconditions with intrinsic properties, or local conventions with universal necessities.

## Partial Tabula Rasa: A Corrective Clarification

The Double Preconditioning Theorem excludes absolute structural neutrality, but it does not deny the existence of partial neutrality. Within any constrained initial condition  $I = P + F$ , the subset  $F$  can be substantial. The framework therefore supports a nuanced position:

- **Absolute tabula rasa** ( $P = \emptyset$ ) is incoherent for describable processes;
- **Partial tabula rasa** ( $F \neq \emptyset$ ) is not only possible but ubiquitous.

Methodologically sound analysis requires identifying, for each domain, which components belong to  $P$  and which to  $F$ , rather than assuming either extreme.

## Practical Use of the Typology

This classification yields a practical analytic checklist:

1. **Explicit Partitioning** – For any proposed model or reform, specify which aspects are treated as predetermined and which as adjustable.
2. **Justification Requirement** – Provide reasons for assigning each element to  $P$  or  $F$ .
3. **Domain Verification** – Confirm that the partition corresponds to the correct level of description.
4. **Error Screening** – Test the proposal against the three error types above.

By making these distinctions explicit, the framework transforms the abstract insight of Double Preconditioning into a concrete methodological tool.

In summary, the recognition that initial conditions contain both predetermined and open components enables a disciplined middle path between two symmetrical illusions: the illusion of total neutrality and the illusion of total fixity. Avoiding these errors—especially their conflation across domains—is essential for coherent modeling, responsible design, and realistic interpretation of any operational process.

## 1.8 Methodology for Distinguishing Predetermined and Free Components

The usefulness of the framework developed in Sections 1.6 and 1.7 depends on a practical question: given an actual system or model, how can an analyst reliably determine which elements belong to the predetermined subset  $P$  and which belong to the relatively free subset  $F$ ? Without such a method, the distinction risks remaining purely rhetorical. This section proposes a structured procedure for making that determination. (For a complete step-by-step audit protocol suitable for practical application, see Appendix B.)

### 1.8.1 Operational Criterion of Necessity

The primary test for classifying an element as predetermined is the *necessity-for-description criterion*:

**Necessity Test:** An element  $x$  belongs to  $P$  if removing or altering  $x$  renders the process undescrivable, non-operational, or incoherent within the chosen modeling framework.

Formally:

$$x \in P \iff \text{Description without } x \text{ is undefined or non-functional} \quad (3)$$

Conversely:

$$x \in F \iff \text{Description remains coherent when } x \text{ is varied} \quad (4)$$

This test grounds the distinction in representational practice rather than intuition.

### 1.8.2 The Substitution Test

A second practical tool is the *substitution test*.

For any candidate element  $x$ :

- If alternative values of  $x$  can be substituted without violating the operational integrity of the model, then  $x$  belongs to  $F$ .
- If no substitution is possible without collapsing the model's coherence, then  $x$  belongs to  $P$ .

**Example:**

- In a neural network, the exact random seed used for initialization is typically  $F$ ;
- The existence of weights, layers, and an update rule is  $P$ .

### 1.8.3 The Counterfactual Robustness Test

Another diagnostic is the *counterfactual robustness test*:

Ask: "Could the process, as currently described, still be meaningfully specified if this element were absent?"

- If the answer is *no*, the element is predetermined.
- If the answer is *yes*, it is a free parameter.

This test deliberately shifts attention from metaphysical necessity to descriptive necessity.

### 1.8.4 Layered Decomposition of Constraints

In many systems, predetermined components operate at multiple levels. We therefore propose a hierarchical decomposition:

1. **Formal constraints** – logical or mathematical structures required for any representation
2. **Architectural constraints** – structural features of the chosen modeling apparatus
3. **Empirical constraints** – inherited features of the target system
4. **Contingent parameters** – tunable or freely specifiable values

The first three layers typically constitute  $P$ ; the fourth constitutes  $F$ .

This layered view prevents the common error of treating empirical constraints as if they were formal necessities.

### 1.8.5 Decision Tree for Classification

For applied analysis, the following decision tree may be used:

1. Is the element required to specify the system at all? – Yes  $\rightarrow P$
2. Is it imposed by the modeling language or medium? – Yes  $\rightarrow P$
3. Is it inherited from the target system and not under analyst control? – Yes  $\rightarrow P$
4. Can it be varied without destroying coherence? – Yes  $\rightarrow F$

This procedure converts the abstract distinction into a repeatable analytic routine.

### 1.8.6 Common Heuristics

Several practical heuristics assist classification:

- **Immutability heuristic:** what cannot be changed within the time-scale of analysis tends to be  $P$ .
- **Replaceability heuristic:** what can be replaced by alternative values or designs tends to be  $F$ .
- **Infrastructure heuristic:** background enabling conditions usually belong to  $P$ .
- **Tuning heuristic:** adjustable hyperparameters typically belong to  $F$ .

These heuristics are fallible, but they provide a useful starting point.

### 1.8.7 Illustrative Applications

Although the present paper remains primarily methodological, brief illustrations clarify the method:

#### **Machine Learning:**

- Loss function, architecture, training algorithm  $\rightarrow P$
- Hyperparameters, dataset sampling, random seeds  $\rightarrow F$

#### **Cognitive Development:**

- Neural architecture, sensory modalities  $\rightarrow P$
- Specific experiences and learned content  $\rightarrow F$

#### **Institutional Design:**

- Legal frameworks, physical infrastructure  $\rightarrow P$
- Policy choices, administrative rules  $\rightarrow F$

These examples demonstrate that the same entity may be  $P$  in one domain and  $F$  in another, reinforcing the domain-relative character of the distinction.

### 1.8.8 Procedural Summary

The methodology for distinguishing  $P$  and  $F$  can therefore be summarized as follows:

1. Explicitly specify the descriptive framework.
2. List all elements presupposed by that framework.
3. Apply the necessity, substitution, and counterfactual tests.
4. Assign elements to  $P$  or  $F$  accordingly.
5. Re-check for domain conflation as described in Section 1.7.

Only after this procedure has been completed can meaningful claims be made about “degrees of freedom” or “starting conditions.” A more detailed ten-step protocol with a standardized audit sheet template is provided in Appendix B.

### 1.8.9 Methodological Payoff

This methodology converts the abstract insight of Double Preconditioning into an operational practice. It prevents both symmetrical errors identified earlier:

- treating fixed constraints as if they were optional, and
- treating adjustable components as if they were fixed.

By forcing analysts to justify every classification, it replaces vague appeals to “beginnings” with explicit accounts of what, exactly, is fixed and what is open.

## Conclusion of Section 1.8

The distinction between predetermined and free components is not a matter of metaphysical speculation but of disciplined representational analysis. The tools outlined above provide a systematic means of making that distinction explicit, repeatable, and defensible within any describable domain.

## 2 Literature Review

This section reviews literature relevant to the formal framework. The empirical findings discussed here illustrate the framework’s applicability; they are not offered as premises or evidence for the core theorem.

### 2.1 Empiricism and Nativism

The debate over innate versus acquired knowledge is among the oldest in philosophy. In *De Anima*, Aristotle described the intellect as comparable to “a writing tablet on which nothing stands written.” The Stoics elaborated this view, with Aetius reporting that for them the mind at birth is “like a sheet of paper ready for writing upon” [Long and Sedley, 1987]. Locke’s *Essay Concerning Human Understanding* [Locke, 1689/1975] systematized this empiricist position, arguing that the mind at birth is devoid of innate ideas and that all knowledge derives from sensation and reflection.

The rationalist counter-tradition, from Plato through Descartes and Leibniz, maintained that certain ideas or structures must be innate. Leibniz’s *New Essays on Human Understanding* [Leibniz, 1765/1996] offered a detailed critique of Locke, arguing that the mind is not a blank tablet but rather like a block of marble whose veins already mark out the figure to be sculpted.

Contemporary nativism, exemplified by Chomsky’s Universal Grammar [Chomsky, 1965, 1980], grounds claims in genetics and neuroscience rather than metaphysics. The “poverty of the stimulus” argument contends that language acquisition cannot be explained by general learning mechanisms alone—children must bring prior structure to the task. This argument has been formalized and debated extensively [Pullum and Scholz, 2002, Clark, 2011, Perfors et al., 2011].

Jerry Fodor’s modularity thesis extended nativist reasoning beyond language, proposing that the mind comprises specialized modules for face recognition, spatial reasoning, number sense, and theory of mind [Fodor, 1983]. Subsequent work has debated the extent and nature of modularity [Carruthers, 2006, Barrett and Barretto, 2011].

The current consensus rejects both pure nativism and pure empiricism: complex traits emerge from dynamic interactions between genetic endowment and environmental input [Sterelny and Griffiths, 1999, Griffiths and Stotz, 2013, Oyama, 2000]. This “interactionist” or “developmental systems” perspective dissolves the nature-nurture dichotomy, emphasizing that development always involves both inherited and environmental resources. Notably, this empirical convergence illustrates (but does not prove) our formal claim about the requirements of describability.

### 2.2 Dynamical Systems and Path Dependence

The mathematical study of dynamical systems provides formal tools for analyzing how initial conditions propagate through time. Poincaré’s foundational work on the three-body problem revealed that deterministic systems could exhibit extreme sensitivity to initial conditions [Poincaré, 1890]. Lorenz’s discovery of deterministic chaos in atmospheric models demonstrated that this sensitivity—the “butterfly effect”—was not merely theoretical but had practical consequences for prediction [Lorenz, 1963, 1993].

The philosophical implications of chaos theory have been extensively analyzed. Kellert argues that chaos fundamentally challenges traditional notions of determinism and predictability

[Kellert, 1993]. Smith examines the epistemic limitations chaos imposes on scientific knowledge [Smith, 1998].

Path dependence in economics and institutional theory extends these insights to social systems. Paul David’s influential analysis of the QWERTY keyboard demonstrated that standards can become entrenched through historical accident rather than optimality [David, 1985]. Brian Arthur formalized increasing returns mechanisms that generate path dependence [Arthur, 1994]. Douglass North applied these concepts to institutional economics, showing how initial institutional choices constrain subsequent development [North, 1990].

The concept has been refined through debates about “lock-in” and the distinction between different forms of path dependence [Liebowitz and Margolis, 1995, Pierson, 2000]. Mahoney provides a systematic typology of path-dependent explanations in historical sociology [Mahoney, 2000].

Recent experimental work by Mittone, Morreale, and Ritala provides direct evidence that initial conditions shape learning trajectories in exploration-exploitation tasks [Mittone et al., 2024].

### 2.3 Inductive Bias in Machine Learning

Machine learning theory provides particularly clear formalization of issues surrounding initial constraints. Mitchell’s foundational definition established that inductive bias comprises any assumptions beyond the training data that a learner uses to generate predictions [Mitchell, 1980, 1997].

The “no free lunch” theorems, proved by Wolpert and Macready, establish formally that no learning algorithm can outperform all others across all possible problems [Wolpert and Macready, 1997]. This result demonstrates that inductive bias is not merely useful but logically unavoidable: without assumptions about the structure of the target domain, learning is impossible.

Geman, Bienenstock, and Doursat analyzed the bias-variance tradeoff, showing how different forms of inductive bias affect generalization [Geman et al., 1992]. This work established that the choice of bias is not arbitrary but must be matched to the problem structure.

Neural network architectures encode inductive biases through their structural choices. Convolutional networks assume spatial locality and translation invariance [LeCun et al., 1998]. Recurrent networks assume sequential dependencies [Hochreiter and Schmidhuber, 1997]. Transformers assume that attention over pairwise relationships is informationally sufficient [Vaswani et al., 2017]. Battaglia et al. provide a comprehensive analysis of relational inductive biases in deep learning [Battaglia et al., 2018].

Lake, Ullman, Tenenbaum, and Gershman argue that human-like learning requires stronger inductive biases than current deep learning systems typically employ [Lake et al., 2017]. Bengio and colleagues examine what inductive biases might be necessary for learning higher-level cognition [Bengio et al., 2022].

### 2.4 Bias in Large Language Models

The emergence of large language models has made questions of inherited structure practically urgent. Bender, Gebru, McMillan-Major, and Shmitchell’s analysis of “stochastic parrots” highlighted how training data biases propagate through model outputs [Bender et al., 2021]. Blodgett, Barocas, Daumé, and Wallach provided a critical survey of how “bias” is conceptualized in NLP research [Blodgett et al., 2020].

Gallegos et al. provide a comprehensive survey distinguishing intrinsic bias (encoded in representations) from extrinsic bias (manifesting in downstream tasks) [Gallegos et al., 2024]. Navigli et al. show that bias enters through data selection choices as early as corpus construction [Navigli et al., 2023].

Bommasani et al.’s analysis of “foundation models” emphasizes that these systems inherit constraints at multiple levels: from architecture, training procedures, data selection, and fine-tuning processes [Bommasani et al., 2021]. Weidinger et al. taxonomize the ethical and social risks arising from these inherited structures [Weidinger et al., 2021].

## 2.5 Embodied and Situated Cognition

The embodied cognition movement challenges the classical view of mind as disembodied symbol manipulation. Varela, Thompson, and Rosch’s *The Embodied Mind* introduced enactivism, arguing that cognition arises through sensorimotor coupling with the environment [Varela et al., 1991]. Clark’s work on extended cognition proposed that cognitive processes can extend beyond the brain into body and environment [Clark, 1998, 2008].

Gallagher developed the enactivist approach further, emphasizing how bodily possibilities shape cognitive processes [Gallagher, 2005, 2017]. Thompson provided a comprehensive philosophical treatment of the mind-body-world relationship [Thompson, 2007].

Barrett and Stout’s recent work examines embodied cognition in light of artificial intelligence, arguing that current AI systems lack the embodied grounding that characterizes biological cognition [Barrett and Stout, 2024]. Chemero’s radical embodied cognitive science pushes the approach further, rejecting mental representations entirely [Chemero, 2009].

Gibson’s ecological psychology, with its emphasis on affordances, provides a complementary perspective on how environmental structure shapes perception and action [Gibson, 1979]. This work has influenced robotics and situated AI [Brooks, 1991].

For our purposes, embodied cognition provides an existence proof: biological cognition is a domain where description of cognitive processes requires reference to non-zero initial structures—bodily, environmental, and developmental.

## 2.6 Philosophy of Models and Representation

The framework’s emphasis on medium constraints connects to extensive work in philosophy of science on the nature of models and representation. Giere’s semantic view of theories emphasizes that models are abstract structures that represent target systems [Giere, 1988, 2004]. Suárez analyzes the inferential conception of representation [Suárez, 2004].

Weisberg distinguishes different types of models—concrete, mathematical, and computational—each with different representational constraints [Weisberg, 2007, 2013]. Frigg and Hartmann provide a comprehensive overview of models in science [Frigg and Hartmann, 2020].

The model-based view of scientific theories raises questions about the relationship between models and reality that parallel our distinction between representational and operational constraints. Cartwright’s work on the “dappled world” emphasizes the local and constrained nature of scientific models [Cartwright, 1999].

Floridi’s philosophy of information provides a framework for understanding the constraints that information structures impose on representation and processing [Floridi, 2011].

## 2.7 Foundations of Mathematics

The paper’s treatment of mathematical emptiness connects to foundational debates about the nature of mathematical objects. The standard set-theoretic foundation builds all mathematics from the empty set via the Zermelo-Fraenkel axioms [Zermelo, 1908, Fraenkel, 1922]. Yet this construction, as we note, presupposes a non-empty axiomatic framework.

Alternative foundations have been proposed. Category theory offers a structuralist alternative where mathematical objects are characterized by their relationships rather than their intrinsic nature [Mac Lane, 1971, Awodey, 2010]. Homotopy type theory provides yet another foundational framework with different primitive assumptions [HoTT, 2013].

Maddy’s philosophical work examines the justification of set-theoretic axioms [Maddy, 1988, 2011]. Shapiro’s structuralism in philosophy of mathematics analyzes what it means for mathematics to study structures [Shapiro, 1997].

These foundational alternatives illustrate our point that mathematical “emptiness” is always relative to a framework. The empty set is not nothing *simpliciter* but the empty set *within* a particular foundational system.

### 3 Formal Framework

#### 3.1 Operational Definition of System

We require a definition that does not presuppose structure, to avoid circularity. We distinguish between “system as ontological entity” and “system as describable process.” Our concern is the latter:

##### OPERATIONAL DEFINITION

A *system* (for purposes of this paper) is anything that can be described as undergoing state transitions.

This definition is minimal: it does not assume structure a priori.

Crucially, structure is not assumed in the definition; it is derived from the minimal requirements for describing transitions. From this definition, we derive (not assume) that the description of systems undergoing change requires non-zero structure. The derivation proceeds through the requirements of description itself.

#### 3.2 Derivation from Requirements of Description

The argument proceeds as follows:

**Step 1:** The description of functional transitions requires at least minimal differentiable states. To describe change, one must distinguish the state before from the state after.

**Step 2:** Description of differentiable states requires a basis for differentiation. A distinction without ground is not a distinction.

**Step 3:** A basis for differentiation constitutes non-empty structure.

**Conclusion:** The description of a process as undergoing change presupposes a non-empty domain of differentiable states.

This derivation is epistemic: it concerns requirements of description, not prior metaphysical commitments about the nature of systems.

#### 3.3 The Principle of Non-Zero Initialization

**Principle 1** (Non-Zero Initialization—Conditional Form). *If a process is to be described as evolving, its description presupposes a non-empty domain of states.*

**Logical form:** *Describability of evolution  $\rightarrow$  non-empty domain*

*This follows from requirements of description, not from metaphysical assumptions.*

This is a claim about the conditions for description and representation, not about the intrinsic nature of reality.

The supporting argument:

- *Premise 1:* A transition is described as change from one state to another.

- *Premise 2:* Describing “one state” and “another state” requires at least two differentiable elements.
- *Premise 3:* An empty domain provides no elements to differentiate.
- *Conclusion:* The description of transitions requires a non-empty domain.

Note: This does not claim empty domains are logically impossible in some absolute sense or that emptiness cannot be represented. It claims that the description of transitions requires something for those transitions to be between.

### 3.4 The Principle of Medium Constraint

**Principle 2** (Medium Constraint). *Modeling frameworks constrain what can be represented, prior to any specific act of representation.*

*This is a claim about expressibility, not about ontology. It concerns representational limits, not metaphysical limits.*

This is a claim about the conditions for description and representation, not about the intrinsic nature of reality.

The argument:

- *Premise 1:* Modeling requires a representational medium with syntax and semantics.
- *Premise 2:* Syntax determines which expressions are well-formed.
- *Conclusion:* Possible models are constrained by the medium before specific models are constructed.

## 4 Double Preconditioning

**Theorem 1** (Double Preconditioning). *For any system  $S$  that is both (1) describable as undergoing transitions and (2) representable in some modeling framework:*

*The initial state is constrained by both:*

- **Internal requirements:** *description requires differentiable states (Principle 1)*
- **External requirements:** *representation requires expressive framework (Principle 2)*

*The space of possible initial states is the intersection of two non-empty constraint spaces.*

*Proof.* Let  $S$  be describable as undergoing transitions, with preconditions  $P$  from Principle 1. Let  $M$  be a modeling framework with constraints  $C$  from Principle 2. The representation of  $S$ ’s initial state in  $M$  is constrained by both  $P$  and  $C$ . Since  $P \neq \emptyset$  and  $C \neq \emptyset$ , the space of representable initial states is doubly bounded.  $\square$

### 4.1 What This Establishes and What It Does Not

This establishes: For systems we can describe, model, and analyze, zero-structure states cannot function as operational starting points for describable processes.

This does **NOT** establish:

- That metaphysical nothingness is impossible in some absolute sense
- That zero-structure states cannot be conceived or represented
- Anything about domains beyond description

The result is strong within its domain—covering science, mathematics, and formal reasoning—but explicitly bounded.

## 5 The Self-Initialization Objection

A sophisticated objection: perhaps structures can emerge during a process rather than pre-exist it. We examine this carefully.

### 5.1 The Objection

Perhaps there exist processes in which the structures required for transition emerge as part of the process itself. If so, structure need not exist prior to the process—it could be self-generated.

This is a serious objection. We examine specific candidates.

### 5.2 Cellular Automata

**Claim:** Complex patterns emerge from simple rules.

**Analysis:** The description of cellular automata requires: (a) grid structure, (b) state space, (c) transition rule, (d) initial configuration. Emergence of patterns occurs within this pre-existing framework.

**Verdict:** Emergence within structured framework, not emergence of framework from nothing. Consistent with our theorem.

### 5.3 Symmetry Breaking

**Claim:** Differentiated states emerge from symmetric initial conditions.

**Analysis:** The description of the symmetric initial state attributes structure to it (the symmetry). The description of the dynamics requires a potential landscape with non-trivial structure.

**Verdict:** Transformation of one structure type into another, not creation from absence of structure. Consistent with our theorem.

### 5.4 Evolutionary Emergence

**Claim:** Evolution produces genuinely novel structures.

**Analysis:** The description of evolution presupposes: replicators, fitness landscape, chemistry. These are substantial prior structures in any description of evolutionary processes.

**Verdict:** Structural transformation, not self-creation from nothing. Consistent with our theorem.

### 5.5 Assessment

In every examined case, apparent “emergence from nothing” is, upon analysis, emergence within a meta-structure that provides the descriptive framework for the emergence.

#### ACKNOWLEDGMENT AND LIMITATIONS

No currently available model demonstrates self-initialization without presupposing a meta-framework for description.

However: Self-bootstrapping processes remain an open foundational question. We do not claim to have proved impossibility in principle—only that no existing describable model achieves it.

The theorem is therefore conditional on current frameworks of description; future conceptual developments could require revision.

## 5.6 What Would Challenge the Theorem

A genuine challenge would require: (a) describable genuine transition, (b) emergence of the framework for transition as part of the process, (c) no presupposed meta-framework for description.

We claim no such process has been described within current frameworks. We do not claim such a process is impossible in some absolute sense for describable processes.

## 6 Mathematical Emptiness

### IMPORTANT DISCLAIMER

This argument does not challenge the legitimacy of the empty set as a formal object.

Mathematical emptiness  $\neq$  Operational emptiness

We are not rejecting standard set theory or claiming mathematics is deficient.

### 6.1 Two Senses of Emptiness

**Representational emptiness:** Formal representation of absence within a symbolic system. “ $\emptyset$ ” represents the set with no members. This is emptiness as object of study.

**Operational emptiness:** Absence of all structure, including absence of any framework for representation. This would be emptiness as condition, not concept.

Our claim: “nothing” can be represented but cannot function as initial condition for describable processes. This is the difference between representing emptiness and beginning from emptiness.

### 6.2 The Empty Set in Formal Systems

Mathematics shows that representations of “nothing” occur within structured formal systems. The empty set exists within ZFC (or alternatives) that have non-empty axiomatic structure: Axiom of Empty Set, Axiom of Pairing, inference rules, language of set theory.

The empty set is an element within a structured theory; it is not a pre-theoretic origin point for describable operations.

The empty set has properties: subset of every set, cardinality zero, uniqueness. These are structural features within set theory. Mathematical emptiness is always relative to a richer system.

### 6.3 The Point

One can represent the empty set precisely because one has a non-empty formal apparatus. The representation of “nothing” is itself something—symbol, concept, object with properties.

This illustrates rather than challenges our claim: representations of emptiness require non-empty representational resources for their description.

## 7 Illustrations from Artificial Intelligence

This section illustrates how the framework applies to AI. These are illustrations of the general theorem’s applicability, not independent evidence or proofs for it.

## 7.1 Inductive Bias

Inductive bias illustrates Principle 1. Biases are prior assumptions determining the hypothesis space. The “no free lunch” theorems establish that bias is unavoidable for any describable learner: no learner outperforms all others across all problems.

A learner with no bias would have no basis for preferring any hypothesis. The requirement of inductive bias for learning illustrates the general requirement of non-empty structure for describable operation.

## 7.2 Architectural Constraints

Neural architectures encode structural assumptions: Convolutional networks assume spatial locality. Recurrent networks assume sequential dependencies. Transformers assume attention suffices.

These are theoretical commitments, not mere implementation details. Architecture provides the structure making describable learning possible.

## 7.3 Large Language Models

LLMs illustrate multiple inheritance levels:

**Architectural:** Human-designed transformer choices about attention, depth, activations.

**Data:** Human-generated training text reflecting human cognition and culture.

**Objective:** Human-specified loss functions defining success.

Whatever generates something imprints constraints on what is generated—a pattern consistent with our general theorem.

## 7.4 Implications

AI produces novelty constrained by inherited structures. This is a strictly operational observation about how these systems function, not a philosophical claim about “true creativity” or the ultimate nature of machine cognition.

“Unbiased AI” is not coherent as stated. The question is which biases are appropriate for which purposes.

# 8 Objections and Boundary Conditions

## 8.1 The Triviality Objection

**Objection:** The claim is tautological.

**Response:** We do not say “systems require structure because they are systems.” We say: “the description of change requires differentiable states.” This is derived from requirements of description, not stipulated by definition. We began with a minimal definition not presupposing structure; structure was derived from the requirements of describing transitions.

## 8.2 The Overreach Objection

**Objection:** The paper claims too much about all systems.

**Response:** We have stated scope conditions explicitly. The claim concerns describable, operational systems. We make no claims about metaphysical nothingness or domains beyond description. Every use of “impossible” in this paper is qualified: impossible as operational starting points for describable processes.

### 8.3 Potential Counter-Models

A genuine counter-model would need: (a) describable genuine transition, (b) emergence of the framework as part of process, (c) no presupposed meta-framework for description.

Some speculative proposals may point toward this: certain topos-theoretic foundations, some quantum gravity interpretations, process philosophies. We do not claim these cannot succeed. We claim none has yet demonstrated the required features while maintaining describability.

### 8.4 Meaningful Approximations

Our principles concern exact zero-structure as operational starting points, not approximation:

**Relative zero-structure:** A system may lack constraints of one type while having constraints of other types. “Untrained model” is meaningful.

**Level-relative descriptions:** At chosen description levels, initial conditions may be negligible.

**Methodological abstraction:** Provisionally treating constraints as absent for investigative purposes.

## 9 Practical Applications

The framework developed in this paper has direct practical applications across multiple domains. The methodology for distinguishing  $P$  and  $F$  components (Section 1.8), combined with the error typology (Section 1.7), provides concrete tools for analysis and design. For historical examples of each error type, see Appendix A; for a systematic audit protocol, see Appendix B.

### 9.1 Machine Learning

Make inductive biases explicit. Match initialization to known domain properties. Document inherited constraints: architecture, initialization, data, objective, evaluation.

### 9.2 Cognitive Science

The framework supports embodied approaches by providing formal grounding for the claim that the description of cognition requires reference to bodily and environmental embedding.

### 9.3 Scientific Modeling

Selecting a modeling framework is a substantive theoretical commitment about which aspects of reality to capture. Treat modeling choices as requiring justification. Recognize that model limits may reflect medium limits.

### 9.4 Institutional Change

“Starting over” and “building from scratch” are structurally misleading as descriptions. Reformers redistribute constraints rather than eliminate them. The question is not “how to start fresh” but “which inherited structures to transform.”

## 10 Conclusion

We have established a conditional theorem: if something is describable as undergoing change, then its description presupposes non-empty structure. This follows from requirements of describability, not from metaphysical stipulation.

Principle 1 (Non-Zero Initialization): the description of transitions requires non-empty domains. Principle 2 (Medium Constraint): representation requires expressive frameworks. Theorem 1 (Double Preconditioning): describable and representable systems have doubly-constrained initial states.

We examined self-initialization candidates—cellular automata, symmetry breaking, evolution—showing each presupposes meta-structure in any description. We distinguished representational emptiness (working with  $\emptyset$ ) from operational emptiness (beginning from zero structure as starting point for describable processes).

The framework applies to domains amenable to scientific study and formal analysis. It does not address questions about metaphysical origins or domains beyond description. The practical utility of the framework is demonstrated through the error typology (Section 1.7), the methodology for distinguishing predetermined from free components (Section 1.8), the extensive illustrations across domains (Appendix A), and the systematic audit protocol (Appendix B).

**The conclusion, precisely stated:** Within describable domains, every operational starting point presupposes non-zero constraints.

This conclusion follows from the logic of description, not from any metaphysical thesis about the nature of being.

#### FINAL STATEMENT OF SCOPE

This paper does not address what may or may not be possible beyond describable domains.

Our claims are domain-restricted theorems about the requirements of describability and operation, not universal metaphysical necessities.

The result is conditional on current frameworks of description; future conceptual developments could require revision.

## A Illustrative Examples of Methodological Errors Across Domains

This appendix provides concrete examples of the three error types identified in Section 1.7, drawn from history, science, policy, psychology, economics, technology, and other domains of human activity. These examples are offered as illustrations, not as evidence for the formal framework. Their purpose is to demonstrate the practical relevance of the distinctions developed in the paper.

### A.1 Error Type I: Misattribution of Freedom (False Neutrality)

This error occurs when predetermined constraints ( $P$ ) are treated as if they were freely modifiable ( $F$ ).

#### History and Politics

- **French Revolution’s “Year One” (1792):** The revolutionary calendar attempted to reset time itself, treating historical and cultural continuity as eliminable. The predetermined weight of religious, agricultural, and social traditions ( $P$ ) was treated as freely replaceable ( $F$ ). The calendar was abandoned within 14 years.
- **Soviet “New Man” ideology:** The assumption that human nature could be entirely reshaped through social engineering, ignoring biological and psychological constraints inherited from evolution.

- **Khmer Rouge “Year Zero” (1975):** The most extreme attempt to treat all prior social structure as eliminable, leading to catastrophic consequences when predetermined economic, educational, and infrastructural requirements reasserted themselves.
- **Post-colonial nation-building:** Borders drawn without regard for ethnic, linguistic, and historical constraints, treating geography and demography as freely specifiable parameters.

## Science and Medicine

- **Radical behaviorism (Watson, Skinner):** The claim that any child could be trained to become “any type of specialist” regardless of innate predispositions, treating genetic and neurological constraints as negligible.
- **Blank-slate theories of language acquisition:** Pre-Chomskyan assumptions that language learning required no innate grammatical structure, ignoring the poverty of the stimulus.
- **Early gene therapy optimism:** Assumptions that genetic diseases could be “simply” corrected by inserting correct genes, underestimating the predetermined complexity of gene regulation networks.
- **One-size-fits-all medical treatments:** Ignoring pharmacogenomic variation by assuming all patients respond identically to standardized dosages.

## Technology and Artificial Intelligence

- **“Unbiased AI” claims:** Marketing language suggesting that algorithms can be free of bias despite predetermined architectural choices, training data distributions, and objective functions.
- **Platform neutrality myths:** Claims that social media platforms are neutral conduits, ignoring how algorithmic ranking, interface design, and business models ( $P$ ) shape discourse.
- **Software rewrites “from scratch”:** The recurring belief that legacy systems can be replaced by clean reimplementations, underestimating embedded institutional knowledge and integration constraints.

## Psychology and Family Relations

- **“Fresh start” illusions in relationships:** The belief that moving to a new city or starting a new relationship eliminates psychological patterns, attachment styles, and behavioral habits formed over decades.
- **Parenting tabula rasa assumptions:** Treating children as infinitely malleable, ignoring temperament, genetic predispositions, and prenatal influences.
- **Ignoring intergenerational trauma:** Assuming families can simply “move on” from historical trauma without addressing inherited psychological and relational patterns.
- **Divorce “clean break” myths:** The assumption that ending a marriage eliminates all prior relational dynamics, especially problematic when children are involved.

## Economics and Policy

- **Shock therapy economic reforms:** Rapid privatization programs assuming markets could function without pre-existing legal, institutional, and cultural infrastructure.
- **Ahistorical development models:** Applying economic models from one context to another without accounting for different institutional inheritances.
- **“Just retrain workers” policies:** Assuming displaced workers can freely transition to new industries, ignoring age, location, social ties, and skill-specificity constraints.

## A.2 Error Type II: Misattribution of Determination (False Fixity)

This error occurs when genuinely modifiable elements ( $F$ ) are treated as if they were predetermined ( $P$ ).

### History and Politics

- **Historical determinism:** Claims that historical outcomes were “inevitable,” ignoring contingent decisions and alternative possibilities that were genuinely available.
- **“End of History” thesis:** The assumption that liberal democracy represented a fixed endpoint, treating political evolution as closed rather than ongoing.
- **Racial and ethnic essentialism:** Treating cultural practices as biologically fixed rather than as modifiable social constructions.
- **Caste system justifications:** Framing social hierarchies as natural and immutable rather than as historically contingent institutions.

### Science and Medicine

- **Genetic determinism:** Overstating the fixity of genetic influence while ignoring epigenetics, gene-environment interactions, and developmental plasticity.
- **“Chemical imbalance” reductionism:** Treating mental health conditions as purely biological, ignoring modifiable psychological, social, and behavioral factors.
- **IQ hereditarianism:** Treating intelligence as entirely fixed by genetics, ignoring educational interventions, nutrition, and environmental enrichment.
- **Neurological fatalism:** The outdated view that adult brains cannot change, prior to discoveries of neuroplasticity.

### Technology and Artificial Intelligence

- **Technological determinism:** Claims that technology follows an inevitable trajectory, ignoring design choices, regulatory options, and alternative development paths.
- **“Algorithms are objective” fallacy:** Treating algorithmic outputs as fixed truths rather than as products of modifiable design decisions.
- **Legacy system fatalism:** Assuming outdated systems cannot be modified when incremental improvements are actually feasible.

### Psychology and Family Relations

- **“People never change” beliefs:** Treating personality and behavior as entirely fixed, ignoring evidence for adult development and therapeutic change.
- **Fixed mindset (Dweck):** The belief that abilities are innate and immutable rather than developable through effort.
- **Learned helplessness generalization:** Extending genuine constraints into domains where agency actually exists.
- **“That’s just how our family is”:** Treating dysfunctional family patterns as immutable traditions rather than modifiable behaviors.

### Economics and Policy

- **TINA (“There Is No Alternative”):** Presenting particular economic policies as the only possible option, foreclosing genuine policy alternatives.
- **Natural rate of unemployment dogma:** Treating a specific unemployment level as structurally fixed when it depends on modifiable policy choices.

- **Poverty as character flaw:** Attributing economic outcomes entirely to individual traits while ignoring modifiable structural factors.

### A.3 Error Type III: Domain Conflation (Category Misplacement)

This error occurs when the  $P/F$  partition from one domain is incorrectly applied to another.

#### Cross-Domain Category Errors

- **Biological → Social:** Treating socially constructed gender roles as biologically determined; using evolutionary psychology to justify contingent social arrangements.
- **Social → Biological:** Lysenkoism’s rejection of genetics in favor of ideologically preferred environmental explanations; denying biological contributions to behavior for political reasons.
- **Mathematical → Empirical:** Treating economic models as if they had the necessity of mathematical theorems; confusing model assumptions with empirical facts.
- **Empirical → Logical:** Treating observed regularities as logical necessities; assuming that because something has always been done a certain way, it must be done that way.
- **Local → Universal:** Treating culturally specific practices as human universals; assuming Western institutional forms are necessary for development.
- **Technical → Political:** Framing political choices as technical necessities (“the algorithm decided”); hiding value judgments behind claims of computational objectivity.
- **Political → Technical:** Assuming that political will can override technical constraints; ignoring physical, computational, or resource limitations.

#### Historical Examples of Domain Conflation

- **Social Darwinism:** Misapplying biological concepts of selection to justify social inequality, conflating evolutionary constraints with political choices.
- **Scientific racism:** Using (flawed) biological measurements to justify social hierarchies, treating political arrangements as if they were natural necessities.
- **Phrenology and criminal justice:** Treating criminal behavior as biologically fixed based on skull measurements, conflating anatomical features with behavioral choices.
- **Eugenics movements:** Conflating genetic inheritance with social worth, and treating social problems as if they required biological solutions.

#### Contemporary Examples

- **Algorithmic sentencing:** Treating recidivism predictions as objective facts rather than as outputs of models encoding particular (modifiable) assumptions about risk.
- **Smart city determinism:** Assuming that data-driven urban management is neutral, ignoring how sensor placement, data categories, and optimization targets embed political choices.
- **Neuroscience in courtrooms:** Using brain scans to argue for diminished responsibility in ways that conflate neural correlates with causal determinism.
- **Economic naturalism:** Treating market outcomes as natural phenomena rather than as results of particular (modifiable) legal and institutional frameworks.

### A.4 Summary Table

### A.5 Methodological Note

The examples collected here are not offered as proofs of the framework but as illustrations of its diagnostic utility. In each case, the error can be precisely characterized using the  $P/F$  notation

Error Type	Structure	Consequence	Paradigm Case
Type I: False Neutrality	$P \rightarrow F$	Ignoring real constraints leads to failed implementations	“Year Zero” ideologies
Type II: False Fixity	$F \rightarrow P$	Foreclosing genuine possibilities leads to fatalism	Genetic determinism
Type III: Do-main Conflation	$(P + F)_{D_1} \rightarrow D_2$	Category errors lead to inappropriate interventions	Social Darwinism

Table 1: Summary of error types with characteristic consequences and paradigm cases.

and the three-fold typology. The framework does not tell us *which* elements belong to  $P$  or  $F$  in any particular domain—that requires domain-specific expertise. What it provides is a structured vocabulary for identifying, classifying, and avoiding these recurrent patterns of misattribution.

## B Pre-Change Constraint Audit Protocol

This appendix provides a systematic protocol for applying the framework developed in this paper to practical situations involving planned changes, interventions, or designs. The protocol operationalizes the  $P/F$  distinction and the error typology from Section 1.7, converting abstract principles into a repeatable audit procedure.

### Step 0 — Define the Target Change

State in one sentence what is being changed (object, scope, success criterion). Specify the level of description (e.g., algorithmic, organizational, cognitive, institutional). If the level is not fixed, explicitly list candidate levels.

### Step 1 — Set the System Boundary

List what is inside the system and what counts as “environment.” Specify the interface: inputs, outputs, and feedback channels. If the boundary is contested, record two boundary options and repeat the protocol for both (domain-sensitivity check).

### Step 2 — Inventory Constraints Before Intervention

Produce a “constraint register” of all constraints plausibly relevant to the change. Include:

- **Internal constraints:** architecture, invariants, conservation-like relations, fixed resources, compatibility requirements
- **Medium constraints:** representation language, measurement apparatus, evaluation metric, dataset, institutional rules, legal/regulatory frame

### Step 3 — Classify Each Item as P, F, or U

For every constraint, assign one label:

- $P$  (predetermined / non-negotiable at the chosen level and horizon)
- $F$  (free / modifiable within the chosen level and horizon)
- $U$  (unknown / undecided pending tests)

**Rule:** Default to  $U$  when uncertain; do not force  $P/F$  prematurely.

## Step 4 — Run the Four Tests on All U Items (and on Any Disputed P/F)

**Necessity Test:** If you remove/negate it, does the system cease to be operational at the chosen level? If yes  $\rightarrow P$ .

**Substitution Test:** Can it be replaced by an alternative without loss of operability? If yes  $\rightarrow$  likely  $F$  (or “ $F$  within a family”). If no  $\rightarrow$  likely  $P$ .

**Counterfactual Robustness Test:** Across plausible counterfactual settings (data, environment, initial states), does it persist as a constraint? If yes  $\rightarrow P$ . If it varies meaningfully  $\rightarrow F$ .

**Medium-Dependence Test:** Does this “constraint” disappear when you change the modeling/measurement medium? If yes  $\rightarrow$  it is a medium constraint ( $C$ ), and you must treat it as  $P$ -within-medium but  $F$  across media (record both).

## Step 5 — Produce the P/F Map and the “Degree-of-Freedom Budget”

Summarize: what is fixed, what is adjustable, and where the unknowns remain. Convert this into a concrete “degrees-of-freedom budget”: list the parameters you are actually allowed to move without breaking operability.

## Step 6 — Identify the Dominant Error Risk (Section 1.7 Taxonomy Gate)

Before acting, explicitly check which methodological failure mode you are most likely to commit:

- **Type I:** treating  $P$  as  $F$  (attempting to redesign invariants)
- **Type II:** treating  $F$  as  $P$  (premature fatalism; refusing feasible change)
- **Type III:** domain confusion (mixing levels of description; importing constraints from one domain/level into another)

If Type III risk is present, stop and redo Steps 0–2 at the corrected level.

## Step 7 — Design Interventions Only Inside F (and Flagged U with Safeguards)

Draft the intervention plan constrained to  $F$  parameters. For any  $U$  you still intend to touch, add a safety clause: a rollback condition and a monitoring signal that detects violation of operability.

## Step 8 — Define Evaluation Without Circularity

Specify success metrics and ensure they are not identical to the medium constraints that generated the model (avoid “metric lock-in”). Record at least one secondary metric that is medium-diverse (different measurement/evaluation lens).

## Step 9 — Pre-Mortem: Simulate Failure as Constraint Violation

Write a short “failure narrative” in formal terms: which  $P$  was mistakenly treated as  $F$ , or which  $F$  was mistakenly treated as  $P$ , or which domain boundary was confused. Convert each narrative into a detection check (what would we observe first?).

## Step 10 — Execute with Monitoring and Update the Constraint Register

Run the intervention. Monitor the operability signals. Update the  $P/F$  map based on observed violations or newly revealed invariants. Treat every unexpected failure as evidence that a hidden  $P$  existed or that the medium imposed an unrecognized constraint.

## Output Artifact — The Constraint Audit Sheet

The protocol produces a one-page summary document containing:

- System boundary and level of description
- Constraint register (complete list)
- $P/F/U$  classification with justifications
- Tests applied and results
- Dominant risk type identified
- Intervention space (parameters in  $F$ )
- Success metrics (primary and secondary)
- Rollback triggers and monitoring signals

## Template: Constraint Audit Sheet

Field	Entry
Target Change	[One-sentence description]
Level of Description	[Algorithmic / Organizational / Cognitive / Institutional / Other]
System Boundary	[Inside: ... / Outside: ... / Interface: ...]
Constraint Register	[List all identified constraints]
P Components	[List with justification]
F Components	[List with justification]
U Components	[List with planned tests]
Tests Applied	[Necessity / Substitution / Counterfactual / Medium-Dependence]
Dominant Risk Type	[Type I / Type II / Type III]
Intervention Space	[Parameters to be modified]
Primary Metric	[Success criterion]
Secondary Metric	[Medium-diverse alternative]
Rollback Triggers	[Conditions for stopping]
Monitoring Signals	[What to observe]

Table 2: Constraint Audit Sheet template for systematic application of the framework.

## References

- Arthur, W. B. (1994). *Increasing Returns and Path Dependence in the Economy*. University of Michigan Press.
- Awodey, S. (2010). *Category Theory*. Oxford University Press, 2nd edition.
- Barrett, H. C. and Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, 113(3):628–647.
- Barrett, L. and Stout, D. (2024). Minds in movement: Embodied cognition in the age of artificial intelligence. *Philosophical Transactions of the Royal Society B*, 379(1911):20230144.
- Battaglia, P. W., Hamrick, J. B., Bapst, V., et al. (2018). Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*.
- Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of FAccT 2021*, pages 610–623.

- Bengio, Y., Lecun, Y., and Hinton, G. (2022). Inductive biases for deep learning of higher-level cognition. *Proceedings of the Royal Society A*, 478(2266):20210068.
- Blodgett, S. L., Barocas, S., Daumé III, H., and Wallach, H. (2020). Language (technology) is power: A critical survey of “bias” in NLP. *Proceedings of ACL 2020*, pages 5454–5476.
- Bommasani, R., Hudson, D. A., Adeli, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47(1–3):139–159.
- Carruthers, P. (2006). *The Architecture of the Mind*. Oxford University Press.
- Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge University Press.
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT Press.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- Chomsky, N. (1980). *Rules and Representations*. Columbia University Press.
- Clark, A. and Lappin, S. (2011). *Linguistic Nativism and the Poverty of the Stimulus*. Wiley-Blackwell.
- Clark, A. and Chalmers, D. (1998). The extended mind. *Analysis*, 58(1):7–19.
- Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press.
- David, P. A. (1985). Clio and the economics of QWERTY. *American Economic Review*, 75(2):332–337.
- Feinman, R. and Lake, B. M. (2018). Learning inductive biases with simple neural networks. *arXiv preprint arXiv:1802.02745*.
- Floridi, L. (2011). *The Philosophy of Information*. Oxford University Press.
- Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press.
- Fraenkel, A. (1922). Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre. *Mathematische Annalen*, 86:230–237.
- Frigg, R. and Hartmann, S. (2020). Models in science. In Zalta, E. N., editor, *Stanford Encyclopedia of Philosophy*.
- Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford University Press.
- Gallagher, S. (2017). *Enactivist Interventions: Rethinking the Mind*. Oxford University Press.
- Gallegos, I. O., Rossi, R. A., Barber, J., et al. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics*, 50(3):1097–1179.
- Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1–58.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin.
- Giere, R. N. (1988). *Explaining Science: A Cognitive Approach*. University of Chicago Press.

- Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science*, 71(5):742–752.
- Griffiths, P. E. and Stotz, K. (2013). *Genetics and Philosophy: An Introduction*. Cambridge University Press.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.
- The Univalent Foundations Program. (2013). *Homotopy Type Theory: Univalent Foundations of Mathematics*. Institute for Advanced Study.
- Kellert, S. H. (1993). *In the Wake of Chaos*. University of Chicago Press.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40:e253.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Leibniz, G. W. (1765/1996). *New Essays on Human Understanding*. Cambridge University Press.
- Liebowitz, S. J. and Margolis, S. E. (1995). Path dependence, lock-in, and history. *Journal of Law, Economics, and Organization*, 11(1):205–226.
- Locke, J. (1689/1975). *An Essay Concerning Human Understanding*. Oxford: Clarendon Press.
- Long, A. A. and Sedley, D. N. (1987). *The Hellenistic Philosophers*, volume 1. Cambridge University Press.
- Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141.
- Lorenz, E. N. (1993). *The Essence of Chaos*. University of Washington Press.
- Mac Lane, S. (1971). *Categories for the Working Mathematician*. Springer.
- Maddy, P. (1988). Believing the axioms I & II. *Journal of Symbolic Logic*, 53(2):481–511, 736–764.
- Maddy, P. (2011). *Defending the Axioms: On the Philosophical Foundations of Set Theory*. Oxford University Press.
- Mahoney, J. (2000). Path dependence in historical sociology. *Theory and Society*, 29(4):507–548.
- Mitchell, T. M. (1980). The need for biases in learning generalizations. *Technical Report CBM-TR-117*, Rutgers University.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Mittone, L., Morreale, A., and Ritala, P. (2024). Initial conditions and path dependence in explorative and exploitative learning. *Technovation*, 129:102895.
- Navigli, R., Conia, S., and Ross, B. (2023). Biases in large language models: Origins, inventory, and discussion. *Journal of Data and Information Quality*, 15(2):1–21.
- North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press.

- Oyama, S. (2000). *The Ontogeny of Information: Developmental Systems and Evolution*. Duke University Press, 2nd edition.
- Perfors, A., Tenenbaum, J. B., and Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, 118(3):306–338.
- Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review*, 94(2):251–267.
- Poincaré, H. (1890). Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, 13:1–270.
- Pullum, G. K. and Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, 19(1–2):9–50.
- Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press.
- Smith, P. (1998). *Explaining Chaos*. Cambridge University Press.
- Sterelny, K. and Griffiths, P. E. (1999). *Sex and Death: An Introduction to Philosophy of Biology*. University of Chicago Press.
- Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science*, 71(5):767–779.
- Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press.
- Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.
- Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30:5998–6008.
- Weidinger, L., Mellor, J., Rauh, M., et al. (2021). Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*.
- Weisberg, M. (2007). Who is a modeler? *British Journal for the Philosophy of Science*, 58(2):207–233.
- Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press.
- Wolpert, D. H. and Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82.
- Zermelo, E. (1908). Untersuchungen über die Grundlagen der Mengenlehre I. *Mathematische Annalen*, 65:261–281.