

Reference-Relative Rationality: A Structural Theory of Model and Regime Revision

Stefan Görig

March 29, 2026

Note on Scope. This paper develops the general formal framework of reference-relative rationality. The framework is intended as the metatheoretical foundation for a family of companion papers (in preparation) that develop domain-specific instantiations — in particular, applications to structured political deliberation, nomological classification systems, and iterative discourse governance. References to these applications appear in the present paper only as illustrations; their systematic development is deferred to the companion work.

1 The Problem of Reference

1.1 The Structural Blind Spot in Accounts of Rational Evaluation

Rational evaluation is commonly analysed as a relation between a model and empirical data. A theory is judged adequate, confirmed, or preferable insofar as its predictions align with observed outcomes. This picture presupposes that “data” are given and that the comparison between theory and world is direct.

This presupposition is structurally false.

Data are not primitive givens. They are outputs of stabilised procedures. They are generated within systems that specify the admissible procedures of observation, the operationalisation of quantities, and the transformations under which meaning is preserved.

These background systems are rarely foregrounded in philosophical accounts. They function as silent infrastructures. Yet every empirical comparison is mediated by them. No theory confronts the world unfiltered; it confronts a reference regime.

The omission of this mediating layer creates a structural blind spot in single-level accounts of rationality.

1.2 Underdetermination Revisited: A Reference-Indexed Phenomenon

Classical underdetermination holds that multiple theories may be empirically equivalent relative to a body of evidence. The usual response invokes extra-empirical virtues—simplicity, coherence, explanatory power. But these discussions often assume that the evidential body is fixed.

In practice, empirical equivalence is indexed to a reference configuration.

A pair of models may be indistinguishable under one regime and discriminable under another. A parameter that is empirically inert at one level of resolution may become identifiable under refined measurement. Conversely, increased tolerance or aggregation may collapse distinctions that were previously discriminative.

Empirical equivalence, therefore, is not absolute. It is reference-relative.

If this is correct, then underdetermination is not merely a property of theories relative to data; it is a property of theories relative to regimes that generate and classify data.

Single-level theories of rational model choice cannot represent this dependency. They lack a formal account of how improvements in resolution or discrimination alter the space of rational preference.

1.3 Reference Regimes as Generative Qualification Structures

A reference regime is not simply a set of observational reports. It is a structured qualification system that determines what counts as a data point, how predictions are admissibly constructed,

how agreement is assessed, and which bridges between theory and observation are legitimate.

Such regimes are not merely abstract specifications. They are realised in instruments, calibration standards, statistical pipelines, and institutional procedures. Their structure is not merely semantic but operational.

Because they specify operating conditions and admissible variability, reference regimes implicitly determine a generative distribution over possible data configurations. They encode assumptions about noise, tolerance, aggregation, and admissible transformation.

Thus, model evaluation is always evaluation relative to a generative background. There is no theory-to-world comparison without an intervening reference architecture.

1.4 Two Interdependent Levels of Rationality

Once the mediating role of reference regimes is made explicit, a structural distinction emerges.

First, there is model-level rationality: Given a stabilised regime, competing models can be evaluated according to adequacy, fit, and disciplined revision constraints.

Second, there is regime-level rationality: Reference regimes themselves can be compared and revised. They differ in discrimination capacity, resolution, robustness against arbitrary weighting and continuity with prior regimes. These two levels are not independent.

Model choice presupposes a regime. Regime evaluation presupposes the discriminability of models. A fully articulated account of rationality must therefore represent the interaction between both levels. Rationality is not exhausted by selecting the best model relative to a fixed background. It includes structured transformation of the background itself.

1.5 Underdetermination

Classical underdetermination is typically formulated relative to a fixed evidential base.

The current framework reformulates underdetermination as reference-indexed: Two theories may be empirically equivalent relative to R while becoming distinguishable under a refined regime R' . This shifts the philosophical question from: "Which theory is true?" to: "Under which reference conditions is discriminative rational comparison possible?"

Underdetermination is not eliminated, but structurally situated.

1.6 The Normative Core Without Metaphysical Inflation

Introducing a second level does not entail relativism, nor does it require a substantive theory of truth.

The framework developed here does not claim that rational procedures guarantee truth. Rather, it formalises the structural constraints under which comparison is non-arbitrary:

- Adequacy is reference-relative.
- Discrimination must not be trivially manipulable.
- Resolution must increase without destroying continuity.
- Revisions must penalise empirically inert extensions.

These constraints are minimal. They do not privilege a metaphysical position. They articulate conditions for disciplined evaluative dynamics within revisable infrastructures.

1.7 Beyond Science: General Scope

The structure extends beyond scientific measurement to any domain where representations compete under stabilised procedures and revisable standards—including metrology, forecasting, institutional decisions, and structured deliberation. Rational comparison depends on the underlying qualification regime.

This work provides a formal architecture for such reference-relative rationality.

1.8 Thesis Statement

The central thesis of this work is the following:

Rational evaluation is structurally reference-relative. Empirical adequacy is not a property of models in isolation, but a relation between a model and a stabilised reference regime that generates and qualifies data.

From this follows a second claim:

Rationality operates on two interdependent levels. At the first level, models are evaluated and revised relative to a given regime. At the second level, reference regimes themselves are subject to structured comparison and revision, governed by discrimination capacity, resolution, robustness, and continuity constraints.

Finally:

Non-arbitrariness in rational practice arises not from fixed foundations, but from disciplined dynamics within revisable reference architectures. Rationality is neither absolute nor unconstrained. It consists in the coherent interaction of adequacy, discrimination, resolution, and constrained revision across both levels.

The remainder of this work develops a formal framework that renders these claims precise.

2 Structural Framework

2.0 Orientation

Section 2 develops the formal architecture of the framework. It proceeds in four stages. Each stage presupposes the preceding one: bridge classes require typed objects; agreement requires bridges; adequacy requires agreement; differentiation requires adequacy; and the generative structure grounds all of the above.

The first stage introduces the basic objects: theories, reference regimes, and their associated spaces. These are the typed entities over which the framework is defined.

The second stage develops the evaluative structure: bridge classes, agreement, projection, and empirical adequacy. These components specify what it means for a theory to be evaluable and adequate relative to a regime.

The third stage addresses differentiation within a regime: empirical inertness, the determinacy argument, integrative burden, and robust misfit. These components specify how theories are distinguished when adequacy alone underdetermines rational preference.

The fourth stage introduces the generative structure of reference regimes: the latent state space, probabilistic kernel, and admissible operating measures. This structure grounds adequacy and misfit in specified data-generating conditions.

The section closes with a statement of commitments: what the framework does and does not impose at this stage.

2.1 Fundamental Sorts

The framework distinguishes the following typed domains:

- \mathcal{T} — a class of theories
- \mathcal{R} — a class of reference regimes
- \mathcal{M} — the class of all model structures associated with theories in \mathcal{T}
- \mathcal{W} — the class of all practical realisations of regimes in \mathcal{R}

For each reference regime $R \in \mathcal{R}$, the following spaces are associated:

- \mathcal{Y}_R — the reference data space
- \mathcal{P}_R — the prediction space
- Θ_R — the latent state space

No comparison occurs across incompatible types. Each theory $T \in \mathcal{T}$ determines a class of models $Mod(T) \subseteq \mathcal{M}$. A model $m \in Mod(T)$ is a structured instantiation of the theory.

At this stage, nothing is said about empirical content. Theories are treated purely structurally.

2.2 Regimes

A reference regime $R \in \mathcal{R}$ is a structured system consisting of:

- A data space \mathcal{Y}_R
- A prediction space \mathcal{P}_R
- An agreement relation $Agree_R \subseteq \mathcal{P}_R \times \mathcal{Y}_R$
- A generative reference model
- A realisation condition

Reference regimes are not epistemically passive containers for data. They actively structure evaluation by specifying how predictions are admissibly formulated, how agreement is assessed, and how data are generated under specified operating conditions. The internal structure of the agreement relation and the generative model are developed in subsequent sections.

2.3 Bridge Classes

Before a theory can be evaluated relative to a regime, a prior condition must be satisfied: admissible bridges between the theory's models and the regime's prediction space must exist.

For each pair (T, R) , we define the set of admissible bridges:

$$\mathcal{G}(T, R) \subseteq \{g : \mathcal{M} \rightarrow \mathcal{P}_R\}$$

A bridge g maps models of T into admissible predictions within the space defined by R . We define:

$$Use(T, R) \iff \mathcal{G}(T, R) \neq \emptyset$$

This condition is a prerequisite for evaluation. A theory that lacks admissible bridges relative to a regime cannot be assessed by that regime. This prevents illicit cross-type comparison and ensures that evaluative claims are always mediated by an explicit interpretive structure.

Bridge classes are part of the regime-theory interface. They are not determined by the theory alone, nor by the regime alone, but by their structural compatibility.

2.3.1 Admissible Bridge Classes

A bridge $g \in \mathcal{G}(T, R)$ connects latent theoretical structure to regime-qualified observables. Not every mapping between Θ_T and \mathcal{Y}_R constitutes an admissible bridge.

A bridge g is admissible relative to regime R iff the following conditions hold:

- **Partition Compatibility:** If two outcomes are equivalent under \sim_R , their pre-images under g must not induce distinctions that violate this equivalence. Bridges must respect the regime’s observational partition structure.
- **Operational Measurability** For all $\pi \in \Pi_R$, the pushforward of the induced distribution via g is well-defined and measurable. In particular, adequacy and discrimination evaluations must be expressible relative to admissible operating measures.
- **Reparametrisation Invariance** If two theoretical representations are related by admissible reparametrisation, admissibility of the bridge is preserved. Bridge admissibility cannot depend on representational artefacts.
- **Non-Triviality** The bridge must connect theoretically relevant distinctions to observational structure. Constant or vacuous mappings are excluded.

Admissibility is therefore a structural constraint imposed by the regime profile. It is not fixed by external metaphysical considerations, but by compatibility with R ’s qualification and operating structure.

Bridge admissibility is regime-indexed and fixed prior to theory comparison; it is not adjusted in response to evaluative outcomes.

The admissibility constraints on bridges mirror the measurement-theoretic insight that links between theoretical magnitudes and observable quantities are mediated by models and procedures rather than fixed by mere reference. Bridge admissibility ensures that this mediation remains structurally accountable within a given regime. In this respect, bridges function analogously to measurement models: they connect latent structure to observable outcomes under specified operating conditions (cf. Tal, 2013).

2.4 Agreement

Given that admissible bridges exist, the agreement relation $Agree_R \subseteq \mathcal{P}_R \times \mathcal{Y}_R$ specifies when a prediction is acceptable relative to a datum under regime R .

The framework permits different structural forms of agreement depending on domain and regime specification. Three canonical cases are particularly relevant.

(i) Tolerance-Based Agreement

$$Agree_R(p, y) \iff d_R(p, y) \leq \epsilon_R$$

where d_R is a regime-specified distance function and $\epsilon_R > 0$ a tolerance parameter. This form underlies measurement contexts in which predictions are acceptable within an error envelope.

(ii) Likelihood-Based Agreement

$$Agree_R(p, y) \iff L_R(y|p) \geq \tau_R$$

where L_R is a regime-specified likelihood function and τ_R a threshold. This form is natural in statistical contexts.

(iii) Loss-Based Agreement

$$Agree_R(p, y) \iff l_R(p, y) \leq \epsilon_R$$

where $l_R : \mathcal{P}_R \times \mathcal{Y}_R \rightarrow \mathbb{R}_{\geq 0}$ is a loss function. Adequacy can equivalently be expressed via bounded expected or worst-case loss.

Regardless of form, $Agree_R$ must satisfy the following minimal structural constraints:

- Well-Typedness: $p \in \mathcal{P}_R, y \in \mathcal{Y}_R$
- Non-Triviality: $\exists p, y : Agree_R(p, y)$
- Regime Fixity: $Agree_R$ is determined by the specification of R , not by the theories under evaluation
- Reparametrisation Invariance: under admissible coordinate transformations preserving regime structure, agreement is preserved

These constraints ensure that the agreement relation functions as a structural component of the regime rather than as an adaptive device tuned to particular theories.

2.5 Projection

Given a theory T and a regime R , the regime-indexed projection is defined as:

$$Pred_R(T) := \{g(m) | m \in Mod(T), g \in \mathcal{G}(T, R)\} \subseteq \mathcal{P}_R$$

This is the set of admissible predictions generated by T under R via all admissible bridges. Projection is regime-indexed because both bridge admissibility and agreement conditions are regime-dependent. It captures the empirically accessible content of T relative to R .

Two structurally distinct theories may share identical projections under a given regime. This possibility motivates the distinction between adequacy and inertness developed in subsections 2.6 and 2.7.

Invariance: Projection is invariant under admissible reparametrisation of models.

2.6 Empirical Adequacy

Empirical adequacy is the central model-level evaluation criterion. It is regime-indexed and bridge-mediated.

Definition: Given $T \in \mathcal{T}$ and $R \in \mathcal{R}$:

$$EA(T, R) \iff Use(T, R) \wedge \forall y \in \mathcal{Y}_R \exists m \in Mod(T), g \in \mathcal{G}(T, R) : Agree_R(g(m), y)$$

This definition has three structural features.

1. *Regime-indexed*: Adequacy depends on \mathcal{Y}_R , the agreement relation, and the bridge structure of R . A theory adequate under a coarse regime may fail under a refined one. Adequacy is stable only relative to fixed reference conditions.
2. *Bridge-mediated*: No direct theory-world comparison exists. Predictions must be admissible within the regime's qualification structure before agreement can be assessed.
3. *Not truth*: Adequacy does not entail that the theory correctly represents unobservable structure. A theory may be adequate relative to R while failing under R' . The framework makes no commitment to truth-conduciveness.

Relation to misfit: The binary notion of adequacy defined above is recovered as the zero-point of a graded misfit measure introduced in subsection 2.11. Specifically: $EA(T, R) \leftrightarrow Fit(T, R) = 0$.

2.7 Empirical Inertness

Projection equivalence is possible without structural identity. A theory T' is empirically inert relative to R with respect to T if:

$$Pred_R(T') = Pred_R(T)$$

Inert extensions do not enlarge the regime-indexed projection space. This means that for all admissible data $y \in \mathcal{Y}_R$, the two theories generate identical predictions relative to R . No data produced under R can discriminate between them.

Inertness does not mean triviality. Additional theoretical structure in T' may be conceptually significant, explanatorily motivated, or mathematically elaborate. The point is purely evaluative: that structure makes no empirical difference relative to the current regime.

Inert extensions are common in scientific practice. The crucial question is how rational revision should treat them. This question is addressed in the following two sections.

2.8 The Determinacy Argument

Empirical inertness creates a structural problem for revision that is independent of any appeal to parsimony or simplicity as epistemic virtues.

Consider model-level revision under a fixed regime R . If inert extensions of a theory incur no structural cost, then any theory may be enriched by arbitrarily many empirically invisible components without affecting either its adequacy or its discriminative standing. The revision operator would then fail to induce a determinate ordering over theories that are projection-equivalent under R . Indeterminacy would proliferate across all equivalence classes defined by projection identity.

This is not a merely technical inconvenience. It means that the revision relation, as defined by empirical fit alone, cannot distinguish between a minimal theory and an arbitrarily inflated variant of it. Any evaluative order induced over such variants would be underdetermined in a sense that no further empirical evidence could resolve — not because the theories are difficult to test, but because the regime lacks the discriminatory resources to differentiate them by construction.

To preserve determinate revision under fixed regime conditions, rational revision must therefore treat projection-equivalent theories as empirically equivalent. Strict preference among them requires a structural cost term independent of empirical fit.

Principle of Determinacy under Fixed Regime: For a fixed reference regime R , rational revision must treat projection-equivalent theories as empirically equivalent. Strict preference among them requires structural cost terms independent of empirical fit.

This principle does not assert that simpler theories are metaphysically superior, nor does it invoke parsimony as an independent virtue. It states only that under a regime whose discriminatory capacity is fixed, empirical indistinguishability forecloses differentiation at the level of adequacy. Penalisation of inert structure is a consequence of maintaining determinacy, not an independent normative commitment.

It follows that if T' extends T by structure that is inert relative to R , then:

$$\text{pen}_R(T') > \text{pen}_R(T)$$

This constraint is local. Under a refined regime R' , previously inert structure may become identifiable and cease to incur penalty. Penalisation is reference-relative, not absolute.

2.9 Integrative Burden

The determinacy argument establishes that inert extensions must be penalised. The integrative burden formalises this requirement.

Let $pen_R(T)$ measure the integrative burden of T relative to R . Intuitively, $pen_R(T)$ increases with theoretical structure that does not project into $Pred_R(T)$.

Formal requirement: pen_R must satisfy the following monotonicity condition:

If T' extends T by adding structure that is inert relative to R , then
 $pen_R(T') > pen_R(T)$.

While the precise implementation of pen_R may vary, its monotonicity under inert extension is essential. This ensures revision does not reward empirically invisible elaboration, preventing underdetermination from degenerating into arbitrariness. pen_R is fixed as part of the regime profile and invariant under admissible reparametrisation.

Two remarks are in order.

First, the monotonicity condition is a constraint on any admissible implementation of pen_R , not a full specification. Different domains may instantiate it differently, provided the constraint is satisfied.

Second, pen_R is regime-relative. Under a refined regime R' , the burden associated with previously inert structure may decrease or vanish if that structure becomes projectively active. Integrative burden tracks identifiability, not intrinsic complexity.

2.10 Generative Structure

The preceding sections defined adequacy and misfit in terms of agreement between predictions and data. To ground these notions in specified data-generating conditions, the regime must be equipped with a generative structure. Each reference regime R includes:

- a latent state space Θ_R
- a probabilistic kernel $K_R(\cdot|\Theta)$ over \mathcal{Y}_R , measurable in Θ
- a class of admissible operating measures $\Pi_R \subseteq Prob(\Theta_R)$

For each $\pi \in \Pi_R$, the induced data-generating distribution is:

$$\mu_{R,\pi}(A) = \int_{\Theta_R} K_R(A|\Theta) d\pi(\Theta)$$

The set $\{\mu_{R,\pi} | \pi \in \Pi_R\}$ represents the admissible data-generating distributions of R . This generative structure ensures that data are not arbitrary inputs but arise under specified operating conditions encoded in the regime.

The explicit inclusion of admissible operating measures is not merely technical. It reflects a central insight from contemporary philosophy of measurement.

Philosophical work on measurement has emphasized that measurement is not a purely representational mapping from pre-given qualitative structure to numerals, but an activity embedded in instruments, procedures, and operating conditions (Chang, 2004; Tal, 2013; see also Tal, 2017).

Treating admissible operating conditions Π_R as an explicit regime component makes this dependence structurally visible. Rather than relegating operating variability to background noise, the present framework incorporates it into the evaluative architecture itself. Regime-relative adequacy and discrimination are therefore explicitly indexed to material and procedural constraints.

Two constraints govern Π_R .

Regime Precommitment: Π_R must be determined independently of the specific theories under comparison. The mapping $(T_i, T_j) \mapsto \text{DiscPair}_R(T_i, T_j, y)$ must not induce modification of Π_R . This prevents selective restriction of operating conditions to favour particular models.

Domain Adequacy: Π_R must reflect the domain for which R is specified. It must be non-empty, operationally feasible, and must cover the declared operating envelope of the regime. It must not be artificially restricted to exclude problematic states solely to increase discrimination.

Invariance: Under bijective reparametrisation ψ of \mathcal{Y}_R , the kernel K_R transforms via pushforward, Π_R transforms accordingly, and $\mu_{R,\pi}$ transforms covariantly. This ensures that discrimination capacity is invariant under mere coordinate change.

2.11 Robust Misfit

Empirical adequacy is binary: a theory either covers all admissible data under R or it does not. For purposes of revision and graded comparison, a continuous measure of misfit is required. Such a measure must be compatible with the regime's generative structure and resistant to manipulation through selective weighting of operating conditions.

Pointwise Agreement Indicator: For a theory T and regime R , define:

$$A_{T,R}(y) = \begin{cases} 1 & \text{if } \exists m \in \text{Mod}(T), g \in \mathcal{G}(T, R) : \text{Agree}_R(g(m), y), \\ 0 & \text{otherwise.} \end{cases}$$

Misfit under π : For an admissible operating measure $\pi \in \Pi_R$:

$$\text{Fit}_{R,\pi}(T) = \int_{\mathcal{Y}_R} (1 - A_{T,R}(y)) d\mu_{R,\pi}$$

This measures the expected proportion of admissible data not covered by T under operating distribution π .

Robust Misfit: Since the choice of a single operating distribution would expose the measure to manipulation, misfit is evaluated against the entire admissible operating envelope:

$$Fit(T, R) = \sup_{\pi \in \Pi_R} Fit_{R,\pi}(T)$$

Robust misfit evaluates a theory against the worst-case admissible operating distribution rather than a single realised one.

Immediate properties:

$$EA(T, R) \leftrightarrow Fit(T, R) = 0$$

If T' is inert relative to T under R , then

$$Fit(T', R) = Fit(T, R)$$

The second property confirms that robust misfit does not discriminate between projection-equivalent theories. Discrimination among such theories requires the structural cost term pen_R , as established in subsections 2.8 and 2.9.

Robust misfit does not introduce additional evaluative standards. It explicates the envelope of admissible operating conditions already encoded in the regime profile via Π_R .

2.12 Commitments

At this stage, the framework commits to the following:

- typed relational structure between theories and regimes
- bridge-mediated evaluation as a necessary condition for adequacy assessment
- regime-indexed projection as the empirically accessible content of a theory
- determinacy of revision through penalisation of inert extensions
- generative grounding of reference regimes in specified operating conditions
- invariance of evaluative notions under admissible reparametrisation

The framework does not yet impose criteria for comparing reference regimes with one another, nor does it specify revision rules that operate across regime transitions. These are introduced in sections 3 through 5.

The evaluative structure developed here is entirely internal to a fixed regime. Its extension to regime-level comparison and coupled revision across both levels is the subject of the remainder of the work.

3 Structural Dimensions of Reference Regimes

3.0 Orientation

Reference regimes are structured objects. Variation in rational assessment may arise not only from changes in theories but from structural variation within regimes themselves. This section isolates three orthogonal dimensions of regime variation.

3.1 Resolution Variation

Beyond discrimination, regimes differ in resolution. Resolution variation changes granularity of admissible distinctions. Define the equivalence relation:

$$y \sim_R y' \Leftrightarrow \forall p \in \mathcal{P}_R : \text{Agree}_R(p, y) \leftrightarrow \text{Agree}_R(p, y')$$

The partition induced by \sim_R reflects the regime's capacity to distinguish data points. A regime R' is finer than R if its induced partition refines that of R . Resolution captures structural refinement independently of probabilistic weighting.

Let $x \in \mathbb{R}$ denote a latent physical state parameter. Consider two theories T_A and T_B where $\alpha, \delta, \omega \in \mathbb{R}$:

$$\begin{aligned} T_A(x) &= ax \\ T_B(x) &= ax + \delta \sin(\omega x) \end{aligned}$$

T_B extends T_A by a microscopic oscillatory term.

Regime R_1 : Coarse Thermometric Measurement

Let $\mathcal{Y}_{R_1} = \mathbb{Z}$ and define agreement by tolerance:

$$\text{Agree}_{R_1}(p, y) \Leftrightarrow |p - y| \leq 1$$

Predictions are implicitly rounded to integer values.

Suppose that for all relevant states x_i :

$$|\delta \sin(\omega x_i)| < 0.5$$

Then rounding yields:

$$\text{round}(T_A(x_i)) = \text{round}(T_B(x_i))$$

.

Hence:

$$\text{Pred}_{R_1}(T_A) = \text{Pred}_{R_1}(T_B)$$

.

The oscillatory term is inert relative to R_1 , and under R_1 , both theories are empirically equivalent.

Regime R_2 : Refined Sensor Measurement

Let $\mathcal{Y}_{R_2} = 0.01\mathbb{Z}$ with agreement:

$$\text{Agree}_{R_2}(p, y) \leftrightarrow |p - y| \leq 0.01$$

.

Suppose now:

$$|\delta \sin(\omega x_i)| > 0.01$$

for a non-negligible subset of states.

Then the oscillatory term becomes empirically identifiable:

$$\text{Pred}_{R_2}(T_A) \neq \text{Pred}_{R_2}(T_B)$$

So, under R_1 , T_B is inert, while under R_2 , T_B is discriminable: empirical equivalence is reference-relative.

3.2 Aggregation Variation

Aggregation can induce projection equivalence without altering latent structure.

Consider price trajectories $p_i(t)$ for goods $i = 1, \dots, n$.

Define two inflation models.

T_A : Fixed-Weight Index

$$\pi_A(t) = \sum_i w_i^0 \frac{p_i(t)}{p_i(0)}$$

.

Weights w_i^0 are fixed at a base year.

T_B : Chain-Weighted Index

$$\pi_B(t) = \prod_{k=i}^t \sum_i w_i^{k-1} \frac{p_i(k)}{p_i(k-1)}$$

.

Weights update dynamically.

T_B contains additional structural responsiveness.

Regime R_1 : Annual Fixed-Basket Measurement

- Annual average prices
- Fixed basket
- Rounding to 0.1%
- Tolerance $\pm 1\%$

Under coarse annual aggregation, substitution effects average out.

For moderate price changes: $\pi_A(t) \approx \pi_B(t)$

after rounding.

Hence: $Pred_{R_1}(T_A) = Pred_{R_1}(T_B)$.

Dynamic reweighting is inert relative to R_1 .

Regime R_2 : Monthly Chain-Weighted Measurement

- Monthly microdata
- Dynamic weight adjustment
- Tolerance $\pm 0.1\%$

Under this regime, substitution and volatility effects become visible.

$Pred_{R_2}(T_A) \neq Pred_{R_2}(T_B)$.

The additional structure in T_B becomes empirically effective.

3.3 Operating Distribution Variation

Reference regimes differ not only in resolution or aggregation, but also in the admissible operating conditions under which data are generated. These conditions are encoded in the set Π_R of admissible operating measures.

Variation in Π_R alters the operating envelope against which theories are evaluated. A theory that performs adequately under laboratory-stabilised conditions may fail under field conditions if the admissible noise structure, sampling frame, or intervention profile changes. Conversely, restricting admissible operating conditions may artificially increase discriminatory power.

Operating distribution variation therefore modifies the comparative structure of theory evaluation without altering resolution or aggregation. It affects which distinctions remain robust across admissible environments.

Resolution, aggregation, and operating distribution are logically independent dimensions of regime structure.

Resolution refinement does not guarantee discrimination increase, since discrimination depends on generative separation across Π_R .

3.4 Structural Independence

Resolution, aggregation, and operating distribution variation are logically independent components of regime structure.

Discrimination concerns the regime's capacity to differentiate between theories.

Resolution concerns the regime's capacity to differentiate between data configurations.

A regime may increase resolution without increasing discrimination (e.g., measuring more precisely but without affecting theoretical distinctions).

Conversely, discrimination may increase if weighting shifts, even without higher resolution.

Rational regime comparison therefore requires both dimensions.

Resolution can increase without altering aggregation: refining the measurement grid of a fixed-basket index leaves its aggregation structure unchanged. Aggregation can change without altering resolution: switching from a fixed-weight to a chain-weighted index at fixed decimal precision. Operating distribution can vary independently of both: extending the admissible noise envelope of an unchanged measurement protocol.

4 Generative Reference Models and Robust Discrimination

4.0 Orientation

The generative structure of reference regimes was introduced in subsection 2.10: each regime R includes a latent state space Θ_R , a probabilistic kernel K_R , and a class of admissible operating measures Π_R . These components specify how data arise under declared operating conditions.

This section builds on that structure to address a distinct question: not whether a theory can match admissible data, but how discriminatively two theories can be distinguished under a regime. Adequacy is a binary, theory-level notion. Discrimination is a comparative, regime-level notion. The two are related but not reducible to one another.

The section proceeds as follows. Subsection 4.1 motivates why discrimination requires robustness against selective weighting of operating conditions. Subsection 4.2 defines robust discrimination capacity formally. Subsection 5.7 establishes invariance of discrimination under reparametrisation.

4.1 Why Robustness is Necessary

Two regimes may share the same data space and agreement relation yet differ substantially in their comparative power. This difference arises from the generative structure: from how data are probabilistically weighted across admissible operating conditions.

If discrimination were evaluated relative to a single arbitrarily chosen distribution μ_R , its capacity could be manipulated by tuning the weighting of operating states. A theory could secure apparent discriminative advantage simply by performing well under a specially constructed distribution that lies outside the declared operating envelope.

To prevent this, discrimination must be evaluated robustly—against the entire admissible operating envelope Π_R , not a single realised distribution within it.

This motivates the infimum construction below: discrimination capacity is the worst-case discriminability across all admissible operating measures.

4.2 Robust Discrimination Capacity

Let $\mathcal{T}_{adm} \subseteq \mathcal{T}$ be a fixed comparison class. \mathcal{T}_{adm} is fixed relative to the domain under consideration. For each $\pi \in \Pi_R$ and theories $T_i, T_j \in \mathcal{T}_{adm}$, the pairwise discrimination under π is:

$$DiscPair_{R,\pi}(T_i, T_j) = \int DiscPair_R(T_i, T_j, y) d\mu_{R,\pi}(y)$$

where $DiscPair_R(T_i, T_j, y)$ indicates whether T_i and T_j are distinguishable at datum y under $Agree_R$, as defined in subsection 2.4.

The robust discrimination capacity of R is:

$$Discrim(R) = \inf_{\pi \in \Pi_R} \frac{1}{|\mathcal{T}_{adm}|^2} \sum_{T_i, T_j \in \mathcal{T}_{adm}} DiscPair_{R, \pi}(T_i, T_j)$$

This quantity reflects the regime’s discriminatory power across the worst-case admissible operating distribution. It cannot be inflated by selective choice of π .

Non-Manipulability: Because $Discrim(R)$ is defined via the infimum over Π_R , a theory cannot secure discriminative advantage by performing well only under operating distributions outside the declared envelope. Selective weighting is neutralised by construction.

Relation to Π_R : The precommitment and domain adequacy constraints on Π_R , established in subsection 2.10, are the structural preconditions for this non-manipulability result. If Π_R were adjustable post hoc or defined relative to the theories under comparison, the infimum construction would not prevent manipulation.

Stability under regime upgrade: If $R \prec R'$, then $\Pi_{R'}$ refines Π_R in the sense established in 5.5. Consequently, discrimination capacity is monotone along upgrade chains: $Discrim(R') \geq Discrim(R)$.

4.3 Invariance and Structural Discipline

Discrimination capacity must be invariant under admissible reparametrisation of the data space. Otherwise, artificial superiority could be generated through mere redescription.

Under bijective reparametrisation ψ of \mathcal{Y}_R :

- K_R transforms via pushforward
- Π_R transforms accordingly
- $\mu_{R, \pi}$ transforms covariantly

It follows that $Discrim(R) = Discrim(R_\psi)$. Regime superiority cannot arise from coordinate change alone.

This invariance condition parallels the reparametrisation invariance of projection (2.5) and adequacy (2.6), and ensures that the evaluative notions at both the model and regime levels are structurally disciplined in the same way.

4.4 Discrimination and Regime Comparison

Robust discrimination capacity is one of the two primary dimensions along which regimes are compared via Dom^* , the other being resolution. The formal upgrade relation is defined in section 5.

The present section establishes only the measurement of discrimination capacity within a fixed regime. Its role in regime-level comparison—and the interaction between discrimination improvement and continuity constraints—is the subject of subsections 5.1 and 5.2.

4.5 Robustness Asymmetry

The apparent asymmetry between the use of the supremum in misfit evaluation and the infimum in discrimination capacity reflects a structural distinction rather than a technical choice.

Misfit is failure-sensitive: a theory is inadequate if there exists an admissible operating condition under which it systematically violates observational constraints. The supremum operationalises this vulnerability.

Discrimination, by contrast, is persistence-sensitive: two states count as robustly distinguishable only if their separation remains stable across the admissible operating envelope. The infimum captures this requirement of cross-conditional stability.

The asymmetry therefore mirrors the logical difference between existential failure and universal persistence, not an arbitrary modelling decision.

This dual structure underlies the *Dom** requirement that discrimination not decrease under refinement while misfit remains failure-sensitive.

5 Regime Upgrade and Structural Continuity

5.0 Orientation

Sections 2 through 4 develop evaluative structure within a fixed reference regime. A theory is assessed for adequacy, inertness, and misfit relative to a given regime; discrimination capacity measures how well that regime separates competing theories. These notions presuppose a stable background.

But reference regimes are not fixed. They evolve: measurement precision increases, admissible noise models change, aggregation procedures are refined. If rational evaluation is reference-relative, then rationality must also govern the transitions between regimes. The question is not whether regimes change, but under what conditions such change constitutes rational upgrade rather than arbitrary replacement.

This section develops criteria for regime-level comparison. It proceeds in five stages.

The first stage establishes when regime comparison is meaningful at all. Regimes are comparable only within a shared domain; without domain constraint, the upgrade relation is undefined (5.1).

The second stage introduces continuity as a structural constraint on regime transitions. An upgrade must preserve evaluative contact with its predecessor on a sufficiently large portion of admissible data (5.2).

The third and fourth stages specify the two dimensions along which a regime must improve: resolution, understood as the fineness of admissible distinctions within the data space (5.3), and robust discrimination capacity, understood as the worst-case separability of competing theories across the admissible operating envelope (5.4).

The fifth stage assembles these conditions into the upgrade relation Dom^* , which formalises what it means for one regime to rationally dominate another (5.5).

The remaining sections address consequences: the possibility of conceptual revolution under continuity constraints (5.2), invariance of the upgrade relation under reparametrisation (5.7), and the partial-order structure induced over admissible regimes within a domain (5.8).

The section thereby establishes the formal conditions under which a regime transition counts as rational improvement rather than arbitrary replacement, and connects these conditions to the model-level evaluative structure developed in sections 2 – 4.

5.1 The Need for Regime-Level Comparison

If empirical adequacy and inertness are reference-relative, then rational comparison of theories depends on the structure of the reference regime.

But reference regimes are not immutable. Historically and structurally, they evolve:

- measurement precision increases,
- admissible noise models change,
- aggregation procedures are refined,
- qualification standards are revised.

A general theory of rationality must therefore provide criteria for comparing reference regimes themselves.

The question is not whether regimes change, but under what conditions such change counts as rational upgrade rather than arbitrary replacement.

5.1.1 Domain Constraint and Regime Comparability

Reference regimes are not universally comparable. Regime comparison is meaningful only relative to a shared domain of application.

Let a domain D consist of:

- a class of target phenomena,
- a data-type structure,

- admissible bridge forms, and
- an operational scope.

We write $R \in \mathcal{R}_D$ to indicate that regime R is specified relative to domain D .

Comparability Condition

Two regimes R and R' are comparable only if:

1. **Data Compatibility:** There exists a partial embedding

$$\phi : \mathcal{Y}_R \rightarrow \mathcal{Y}_{R'}$$

or vice versa.

2. **Prediction Compatibility:** Their prediction spaces admit non-trivial overlap.
3. **Bridge Compatibility:** There exists at least one theory $T \in \mathcal{T}_{adm}$ such that

$$\mathcal{G}(T, R) \neq \emptyset \text{ and } \mathcal{G}(T, R') \neq \emptyset$$

If these conditions fail, the relation $Dom^*(R, R')$ is undefined.

Anti-Domain-Shrinkage Condition A regime R' must not count as an upgrade of R if discrimination improves solely because the domain has been artificially restricted.

Formally, if

$$\mathcal{Y}_{R'} \subset \mathcal{Y}_R$$

and discrimination increases only because problematic data are excluded, then $Dom^*(R, R')$ does not hold.

Regime refinement must preserve domain coverage in the sense that previously admissible data remain interpretable under the refined regime.

Local Partial Orders

Within a fixed domain D , the relation

$$\prec_D$$

defined via Dom^* induces a partial order over \mathcal{R}_D .

There is no claim that:

- regimes across distinct domains are comparable, or
- that a global maximal regime exists.

Rational regime comparison is domain-local.

Reference to the Running Examples

In the temperature case (3.1), both R_1 and R_2 belong to the same thermometric domain. The refined regime increases resolution without altering the target phenomenon.

In the inflation case (3.2), both regimes operate within the same macroeconomic aggregation domain. Monthly refinement does not redefine the target variable but increases discriminatory granularity.

By contrast, thermometric regimes and inflation regimes are not mutually comparable, as their data spaces and bridge structures are disjoint.

5.2 Continuity as a Structural Constraint

A regime upgrade cannot sever all connection to its predecessor without collapsing evaluative continuity.

Let R and R' be reference regimes.

We define a partial transport map:

$$\phi_{R \rightarrow R'} : \mathcal{Y}_R \rightarrow \mathcal{Y}_{R'}$$

.

This map need not be total. However, it must preserve evaluative continuity on a large portion of admissible data.

Formally, for some small $\eta \geq 0$:

$$\inf_{\pi \in \Pi_R} \mu_{R,\pi}(Dom(\phi_{R \rightarrow R'})) \geq 1 - \eta$$

.

This "fast continuity" condition ensures that most admissible data under R remain interpretable under R' .

Revolutionary discontinuity is allowed only in controlled measure.

The parameter η formalises permissible deviation from strict domain preservation. It does not encode substantive preference but reflects the practical fact that regime refinement may involve minor boundary adjustments. In idealised settings η may be set to 0, but it must be sufficiently small to preserve discriminatory continuity in the declared domain.

η is fixed as part of the regime profile and not chosen post hoc.

In practice, η is determined by the declared operational scope of the predecessor regime: it reflects the proportion of previously admissible data configurations that the new regime explicitly acknowledges as boundary cases. It is not a free parameter available for post hoc adjustment, but a commitment encoded in the transition specification.

5.3 Resolution Refinement

As introduced previously, each regime induces a partition over its data space.

A regime R' has higher resolution than R if:

$$\forall y, y' : y \sim_{R'} y' \rightarrow y \sim_R y'$$

.

Thus, R' distinguishes at least as finely as R .

Resolution refinement represents increased structural granularity without loss of comparability.

5.4 Robust Discrimination Improvement

Beyond resolution, regime comparison must consider discrimination capacity.

A regime R' improves upon R if:

$$Discrim(R') \geq Discrim(R)$$

,

with strict inequality in at least one relevant dimension.

Because $Discrim$ is robust with respect to Π_R and $Pi_{R'}$, this comparison cannot be manipulated by selective weighting.

5.5 Dom*

We define the regime upgrade relation Dom^* as follows.

$Dom^*(R, R')$ holds if:

- R' satisfies fast continuity relative to R .
- R' has resolution at least as fine as R .
- $\Pi_{R'}$ refines Π_R in the sense that admissible operating measures under R' remain compatible with those declared under R and do not introduce discriminatory advantage by selective restriction

- R' has discrimination capacity at least as great as R .
- R' is realisable ($\exists w'$ such that $Realize(R', w')$).

At least one of the resolution or discrimination conditions is strict.

We then define:

$$R \prec R' \leftrightarrow Dom^*(R, R')$$

This relation formalises rational regime upgrade.

Π -Refinement

We say that $\Pi_{R'}$ refines Π_R up to $\eta(R)$ iff:

For every admissible operating measure $\pi' \in \Pi_{R'}$ there exists a legacy measure $\pi \in \Pi_R$ such that, under the transport map $\phi_{R \rightarrow R'}$,

1. the transported legacy domain retains at least $1 - \eta(R)$ of its probabilistic mass, and
2. measurable distinctions recognised under π remain approximately preserved under π .

In particular, refinement must not be achieved by selectively excluding operating conditions that carried non-negligible mass under Π_R , except within the tolerated bound $\eta(R)$. Discriminatory gains obtained through systematic restriction of previously admissible operating environments do not constitute refinement.

5.6 Revolution Without Arbitrariness

Continuity constrains arbitrary change, not conceptual evolution. R' may alter bridges, prediction spaces, and assumptions if it preserves structural continuity on sufficient prior data.

5.7 Invariance and Structural Discipline

The Dom^* relation is invariant under reparametrisation of data spaces, provided that:

- generative kernels transform via pushforward,
- admissible operating measures transform covariantly,
- and partition structures are preserved.

This prevents artificial upgrades through mere redescription.

5.8 Regime Hierarchies and Partial Order

The relation \prec induces a partial order over admissible regimes within a domain. It need not produce a total ordering. Competing upgrade paths may coexist. The framework therefore supports regime hierarchies rather than a single linear progression.

Rationality does not require convergence to a unique final regime.

6 The Chemical Revolution as Structured Regime Upgrade

6.1 Historical Delimitation

The following case study reconstructs a well-known episode in late eighteenth-century chemistry: the transition from phlogiston-based accounts of combustion and calcination to oxygen-based, conservation-oriented chemistry associated with Lavoisier.

The reconstruction is deliberately restricted to the domain of laboratory combustion phenomena, in particular the transformation of metals into calxes (oxides). The aim is not to provide a comprehensive historical narrative, but to demonstrate how the framework models a structured regime transition within a fixed domain of phenomena.

The domain D throughout this section is:

Combustion and calcination of metallic substances under laboratory conditions.

Although the Chemical Revolution is a canonical case in the philosophy of science, it is used here not as a historiographic battleground but as a controlled demonstration of how operating envelopes, bridge admissibility, and regime upgrades can be reconstructed within a fixed domain. The aim is structural illustration rather than historical reinterpretation.

6.2 Regime R : Phlogiston Reconstruction

We reconstruct the phlogiston framework as a reference regime R in the sense of section 2.

Domain

D as specified above.

Latent State Space Θ_R

The latent state space includes:

$$\Theta_R = \{metal(M), calx(M), airstate, \phi(M)\}$$

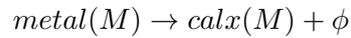
,

where $\phi(M)$ denotes the phlogiston content attributed to a substance.

Phlogiston is treated as a latent magnitude without strict quantitative conservation constraints.

Qualification Structure K_R

The admissible transformation schema includes:



Air plays a background role, but no explicit mass conservation principle is imposed as a regime constraint.

Observable Outcome Space \mathcal{Y}_R

Observations include:

- qualitative classification (metallic / calx),
- visible combustion phenomena,
- coarse mass comparisons.

The induced partition \sim_R identifies outcomes as equivalent whenever they coincide at the level of qualitative classification.

Resolution is therefore comparatively coarse.

Operating Envelope Π_R

Admissible operating measures include:

- open vessels,
- uncontrolled air composition,
- limited precision of balances,
- variable ambient conditions.

Quantitative mass balance is not structurally enforced as a regime requirement.

Agreement Relation

A model m satisfies $Agree_R(m, y)$ if:

- the qualitative transformation is correctly represented,
- no explicit contradiction of observable outcomes arises.

Weight increase in calcination does not, by itself, violate agreement under R .

Discrimination

Discrimination under R is limited to qualitative contrasts.

States differing only in quantitative mass relations remain empirically indistinguishable within Π_R .

Thus $Discrim(R)$ is restricted by coarse resolution and operating conditions.

Structural Burden

Ad hoc hypotheses (e.g., negative phlogiston mass or auxiliary absorption assumptions) increase pen_R , but remain admissible within the regime so long as qualitative agreement is preserved.

6.3 Regime R' : Oxygen and Conservation

We now reconstruct the oxygen framework as regime R' over the same domain D .

Latent State Space $\Theta_{R'}$

$$\Theta_{R'} = \{metal(M), oxide(M), oxygen\ fraction, air\ composition\}$$

Total mass is treated as a conserved scalar quantity:

$$m_{total} = constant$$

Qualification Structure $K_{R'}$

Admissible transformations include:



subject to quantitative mass conservation constraints.

Observable Outcome Space $\mathcal{Y}_{R'}$

Observations include:

- precise mass measurements,
- gas volumes,
- controlled air intake.

The induced partition $\sim R'$ refines $\sim R$:

outcomes are equivalent only if quantitative mass relations agree within specified tolerance.

Resolution is strictly finer.

Operating Envelope $\Pi_{R'}$

Admissible operating measures include:

- closed vessels,
- calibrated balances,
- controlled atmospheric composition,
- repeatable quantitative protocols.

The operating envelope is more tightly specified, but does not exclude previously admissible qualitative situations beyond bounded tolerance.

Agreement Relation

Agreement under R' requires:

- qualitative correctness,
- quantitative mass conservation within tolerance.

Violation of mass balance constitutes regime-level disagreement.

Discrimination

States differing in oxygen uptake become distinguishable.

Quantitative mass shifts generate stable discrimination across admissible $\Pi_{R'}$.

Hence:

$$Discrim(R') > Discrim(R)$$

Structural Burden

Models invoking phlogiston-like latent magnitudes incur high burden under R' , as they require systematic auxiliary adjustments to maintain conservation constraints.

6.4 Verification of $Dom^*(R, R')$

We now verify the Dom^* conditions from section 5 .

Domain Constraint

Both regimes concern the same domain D . No contraction of subject matter occurs.

Fast Continuity

Let $\phi_{R \rightarrow R'}$ map qualitative combustion events to their quantitative counterparts. Qualitative observations remain interpretable under R' . For admissible $\pi \in \Pi_R$, the transported mass of interpretable cases exceeds $1 - \eta(R)$ for small $\eta(R)$. The vast majority of qualitative combustion classifications remain stable under quantitative reinterpretation; hence $\eta(R)$ is small in practice.

Fast continuity holds.

Resolution Refinement

The partition induced by $\sim_{R'}$ strictly refines that of \sim_R . Quantitative mass distinctions subdivide qualitative classes.

Discrimination Non-Decrease

Under the refined operating envelope, states differing in oxygen uptake become robustly distinguishable - discrimination increases. The increase in discrimination arises from conservation constraints interacting with $\Pi_{R'}$, not merely from finer measurement resolution.

Π -Refinement

The shift to closed-vessel protocols does not eliminate non-negligible classes of previously admissible combustion events beyond tolerated $\eta(R)$. Improved discrimination is achieved through controlled measurement, not through selective exclusion of inconvenient operating conditions.

All Dom^* conditions are satisfied - thus: $R \prec R'$.

6.5 Model-Level Consequences

Within regime R , competing phlogiston models achieve comparable fit at the qualitative level. Differences arise primarily in structural burden.

Within regime R' , the oxygen theory minimizes robust misfit relative to $\Pi_{R'}$, satisfies conservation constraints without auxiliary hypotheses and incurs lower structural burden.

Under lexicographic revision (subsection 7.1), the oxygen theory is selected within R' .

Importantly, this selection occurs after regime transition; it is not representable as mere parameter adjustment within R .

6.6 Structural Interpretation

This reconstruction illustrates that regime change can involve refinement rather than incommensurability, that increased discrimination need not rely on metaphysical commitments, and that continuity and structural transformation can coexist.

Rationality remains reference-relative yet structurally disciplined. Conceptual reinterpretation occurs at the level of latent structure, but the evaluative architecture preserves continuity of observable classifications.

The episode exemplifies a structured regime upgrade rather than an arbitrary replacement.

7 Rational Revision on Two Interacting Levels

7.0 Orientation

We have defined reference-relative adequacy, empirical inertness and integrative burden, generative reference regimes, robust discrimination and resolution, and regime-level upgrade via Dom^* . These components establish evaluative criteria. Rationality, however, is dynamic. A general theory must explain how revisions occur across both levels in a disciplined manner.

7.1 Model-Level Revision

Empirical evaluation and structural discipline operate at the model level relative to a fixed reference regime R . Revision proceeds in three ordered stages.

Stage 1: Empirical Discipline

Select theories T' that minimize robust misfit:

$$T^{(1)} \in \arg \min_{T'} Fit(T', R)$$

.

If exact minimizers do not exist, ϵ -minimizers may be used.

This stage ensures maximal conformity with the regime-indexed admissible operating envelope.

Stage 2: Structural Discipline

Among the theories selected at Stage 1, minimize regime-indexed structural burden:

$$T^{(2)} \in \arg \min_{T' \in T^{(1)}} pen_R(T')$$

.

Structural burden measures empirical inertness and identifiability constraints relative to the regime.

This stage prevents inert enrichment without empirical gain.

Stage 3: Stability Constraint

Among the theories selected at Stage 2, minimize structural distance from the prior theory:

$$T^* \in \arg \min_{T' \in T^{(2)}} d_T(T, T')$$

This stage stabilizes revision and avoids unnecessary theoretical displacement.

Lexicographic Structure

The ordering is lexicographic:

1. Empirical conformity has priority over structural economy.
2. Structural economy has priority over minimal displacement.
3. No weighted trade-off is permitted within a fixed regime.

Trade-offs between empirical adequacy and structural burden arise only through regime revision, not through model adjustment.

The lexicographic ordering might seem unrealistically strict. In practice, researchers do trade empirical fit against structural parsimony. The framework does not deny this. Rather, it claims that such trade-offs, when rational, are expressible as regime-level decisions: choosing a regime under which the preferred theory is not inert is not a weighted compromise within a fixed regime but a structured transition to one with greater discriminatory capacity. What appears as a trade-off within a fixed regime is, on closer analysis, an implicit regime shift.

7.2 Regime-Level Revision

When model-level revision stagnates—e.g., when discrimination is low or inert structures accumulate—the rational response may lie not in further model adjustment, but in regime refinement.

Regime revision is governed by Dom^* :

$R \prec R' \Leftrightarrow R'$ satisfies continuity, resolution refinement, discrimination improvement, and realisability.

Thus, regime revision is not arbitrary replacement but structured upgrade.

7.3 Coupling Between Levels

The two revision processes are interdependent.

Upward influence:

If model revision repeatedly encounters inert equivalence classes under R , this may signal insufficient resolution or discrimination.

This creates pressure for regime upgrade.

Downward influence:

If $R \prec R'$, then:

- some previously inert distinctions may become identifiable,
- $Fit(T, R')$ may change,
- pen'_R may differ from pen_R .

Thus, regime refinement reshapes the model landscape.

Rational dynamics therefore consist of alternating or nested revision processes.

8 Grounding, Regress, Relativism, Literature

8.0 Orientation

The framework developed here extends structuralist approaches to scientific theories by making explicit what is often presupposed: that empirical evaluation always occurs relative to a structured reference regime.

Traditional structuralist accounts focus on:

- model classes,
- empirical substructures,
- and theory-data relations.

The present framework retains this orientation but introduces an additional layer: the generative and evaluative structure of reference regimes themselves.

In this sense, it does not replace structuralism but generalises it by internalising the conditions under which empirical comparison becomes possible.

8.1 Regress

A potential objection concerns the status of reference regimes themselves. If theories are evaluated only relative to regimes, and regimes are revisable, does the framework require a further meta-regime to evaluate regimes? If so, one might suspect an infinite hierarchy of meta-level standards.

The framework does not entail such a hierarchy.

First, regime appraisal is domain-local. For a fixed domain D , regimes $R \in \mathcal{R}_D$ are compared via the upgrade relation \prec_D (Def. 5.5). A regime R' qualifies as an improvement over R only if it satisfies:

- continuity relative to R
- resolution refinement,
- operating-envelope refinement ($\Pi_{R'}$ refines Π_R)
- increased or preserved discriminatory capacity, and
- realisability.

These criteria are internal to the domain. No additional meta-domain is required to apply them.

Second, justification of a regime is not construed as external certification. At any time, evaluation is anchored in a stabilised regime R whose generative structure, operating envelope, and agreement relation are precommitted. Critique of R proceeds by constructing a candidate R' , within the same domain and testing whether $Dom^*(R, R')$ holds. Revision is therefore expressed as transition within a partially ordered set (\mathcal{R}_D, \prec_D) , not as appeal to a higher evaluative layer.

Third, so-called meta-evaluations can be represented as domain shifts rather than hierarchical ascent. For example, auditing whether Π_R was properly precommitted belongs to a methodological or governance domain distinct from the empirical domain of R . Such shifts do not generate an infinite regress, since comparability is explicitly restricted by the domain constraint (§5.1).

We may summarise this as follows.

Principle (Local Closure of Regime Evaluation): For any domain D , regime appraisal is closed under the upgrade relation \prec_D . No additional meta-regime is required to justify transitions within \mathcal{R}_D .

The framework therefore rejects both absolute foundationalism and infinite regress. Regime rationality is stabilised locally through structured revision, not through appeal to a hierarchy of higher-order evaluative standards.

8.2 Relativism

The Dom^* criteria do not function as substantive epistemic virtues. They do not privilege explanatory depth, simplicity, predictive scope, or metaphysical realism.

Rather, they articulate minimal structural constraints required for non-arbitrary regime comparison.

Each condition in Dom^* blocks a distinct form of evaluative manipulation: coarse partitioning, semantic rupture, and selective exclusion of operating environments. Together they secure invariance of evaluative structure under regime change without imposing substantive epistemic goals.

It does not assert that later regimes are truer or closer to reality; it asserts only that regime transitions can be assessed without collapsing into opportunism.

In this sense, the normative force of Dom^* is minimal: it consists in preserving the possibility of structured comparison rather than imposing substantive epistemic goals.

8.3 Structuralism

Structuralist approaches (Suppe1979, Sneed, Balzer) model theories as families of structured models connected to empirical substructures via interpretation functions. The present framework extends this by introducing a second structural layer: the reference regime as a formally specified system governing data generation and qualification. Where structuralism focuses on theory-model relations, we formalise the evaluative infrastructure itself—admissible bridge classes, agreement relations, generative kernels, operating measures, and regime-level comparison.

Van Fraassen characterises empirical adequacy as embedding observable phenomena into theory models without commitment to unobservable structure. We share this anti-metaphysical restraint but differ in treating observational structure: rather than presupposing stable observables, we make explicit the reference regime determining which data are admissible, how agreement is assessed, and under which conditions evaluation occurs. Empirical adequacy is therefore regime-relative, not merely model-relative. The framework internalises into formal structure what constructive empiricism treats as background practice.

The difference is not merely one of emphasis. Van Fraassen’s account presupposes that the class of observable phenomena is determinate antecedently to theory choice. The present framework denies this presupposition: what counts as an observable datum is itself regime-constituted. Empirical adequacy, on the constructive empiricist account, is a relation between a theory and a pre-given empirical world; here, it is a three-place relation between a theory, a data point, and a reference regime. This is not a minor extension but a structural revision of the adequacy concept.

8.4 Bayesianism

Bayesian approaches evaluate models through posterior updating based on priors and likelihood functions. Penalisation of complexity arises through marginal likelihood or via information criteria such as AIC or BIC.

The present account is compatible with Bayesian model comparison at the model level. However, it differs in two structural respects.

First, Bayesian updating presupposes a fixed likelihood structure. In contrast, the present framework treats the generative structure (Θ_R, K_R, Π_R) as part of the reference regime and therefore itself subject to structured revision.

Second, penalisation of empirically inert structure is derived from regime-relative identifiability constraints rather than from prior probability assignments. The role of pen_R is not to encode simplicity as an independent virtue, but to preserve determinacy of revision under fixed discriminatory capacity.

In this sense, the framework can be interpreted as providing structural preconditions for Bayesian updating rather than as competing with it.

A genuine tension remains. Bayesian approaches ground penalisation of complexity in prior probability assignments, which are themselves subject to coherence constraints but not to regime-indexed identifiability. The present framework treats pen_R as derived from discriminatory capacity rather than from prior belief. These two foundations are not equivalent: a Bayesian might assign non-negligible prior mass to inert theoretical structure, while the present account requires such structure to incur positive integrative burden regardless of prior. The accounts converge only when Bayesian priors are themselves calibrated to regime-relative identifiability – a substantive constraint that standard Bayesianism does not impose.

8.5 Regime Structure and Measurement Theory

Recent work in the philosophy of measurement has emphasized that measurement standards and procedures are revisable and theory-dependent. Chang’s analysis of thermometry argues that reliable measurement can emerge through iterative stabilization of practices without presupposing a fixed external standard (Chang, 2004). Tal’s model-based account similarly stresses that measurement warrants depend on instruments, calibration procedures, and operating conditions rather than on a direct mapping from theory to world (Tal, 2013).

These contributions articulate the non-neutrality and revisability of measurement regimes. However, they remain largely descriptive or historically reconstructive. The present framework complements this literature by providing a formal architecture in which such revisability yields a principled form of reference-relativity. Empirical equivalence becomes indexed to regime structure, and regime change is modelled as a constrained upgrade relation (Dom^*) preserving interpretability while increasing robust discrimination.

In this sense, the framework does not compete with measurement theory; it systematizes its structural implications. Where Chang emphasizes epistemic iteration and Tal emphasizes model-mediated measurement, Dom^* provides formal conditions under which regime transitions can be evaluated as non-arbitrary without appeal to global epistemic virtues or metaphysical convergence claims.

8.6 Delimitation

This framework applies to practices where models and regimes are evaluated relative to stabilised qualification systems and revisable conditions.

It formalises a middle position between cumulative refinement and revolutionary rupture: regime upgrades permit conceptual change under structured continuity constraints, enabling coexisting partial hierarchies without convergence guarantees.

The normative core is minimal—requiring reference-relative adequacy, discrimination capacity, resolution refinement, non-triviality, reparametrisation invariance, and disciplined revision—providing

structural constraints for non-arbitrary comparison rather than substantive epistemic virtues.

Empirical inertness explains why indistinguishable theories cannot be rationally discriminated at a given regime level: the integrative burden principle follows from identifiability constraints, making empirically inert elaborations rational improvements only when regime refinement renders them identifiable.

The central contribution is making explicit the reference structures underwriting rational comparison: integrating model-level and regime-level evaluation clarifies how empirical adequacy depends on structured generative conditions, how underdetermination is regime-indexed, and how evaluative infrastructures evolve without arbitrariness.

Rationality is not a fixed foundation but a structured dynamic within revisable reference architectures. These are structural constraints on comparison, not metaphysical theses about truth or convergence.

By 'minimal' we mean that the framework prescribes no global epistemic virtues, assumes no convergence to truth, and guarantees no long-run success—it specifies only those structural conditions necessary to ensure that comparison within a dominated regime remains non-arbitrary.

A Technical Conditions and Formal Guarantees

A.1 Domain Assumptions

We assume throughout:

- \mathcal{T}_{adm} is non-empty.
- $\forall R \in \mathcal{R} : \Pi_R \neq \emptyset$.
- For all theories $T \in \mathcal{T}_{adm}$ and all regimes R under active comparison: $\mathcal{G}(T, R) \neq \emptyset$.
- $\forall R \exists w : Realize(R, w)$.

This does not contradict subsection 2.3. The Use-condition remains a non-trivial prerequisite for evaluation. The present assumption restricts attention to those theory-regime pairs that satisfy it; pairs that do not are excluded from the admissible comparison class by definition.

These conditions prevent trivial degeneracy.

A.2 Measurability and Integrability

- $\forall R : (\mathcal{Y}_R, \Sigma_R)$ is a measurable space.
- $\forall \Theta \in \Theta_R : K_R(\cdot | \Theta)$ is a probability kernel measurable in Θ .
- $\forall \pi \in \Pi_R : \mu_{R, \pi}$ is well-defined.

- $DiscPair_R(T_i, T_j, y)$ is assumed measurable in y .

These assumptions guarantee that:

$DiscPair_{R,\pi}(T_i, T_j)$ is well-defined,

$Discrim(R)$ exists in $[0, 1]$.

A.3 Existence Conditions for Lexicographic Revision

A.3.1 Assumptions on the Theory Space

Assume:

1. The admissible theory space \mathcal{T}_{adm} is σ -compact or locally compact.
2. The functions
 - $Fit(T, R)$,
 - $pen_R(T)$,
 - $dT(T_0, T)$ are lower semi-continuous.
3. Sublevel sets of these functions are closed.

These assumptions are standard in structural model-selection contexts.

A.3.2 Existence of Stage-1 Minimizers

Under the above assumptions, the set

$$\mathcal{T}^{(1)} = \arg \min_{T'} Fit(T', R)$$

is non-empty or admits ϵ -minimizers.

A.3.3 Existence of Stage-2 Minimizers

Restricting to $\mathcal{T}^{(1)}$, the structural burden function

$$pen_R(T)$$

attains a minimum on this set under the same compactness and lower semi-continuity assumptions.

A.3.4 Existence of Stage-3 Minimizers

Restricting to $\mathcal{T}^{(2)}$, the distance function

$$dT(T_0, T)$$

admits minimizers or ϵ -minimizers.

A.3.5 Remarks

Lexicographic revision therefore reduces to iterated minimization under standard compactness assumptions. No weighted objective function is required.

A.3.6 Remark on Domain Applicability

The compactness and lower semi-continuity assumptions invoked in A.3.1 – A.3.4 are standard in structural model-selection contexts. Their satisfaction is not trivial, however, in applications to structured deliberation and political discourse. In such domains, the theory space T_{adm} may lack natural compactness properties, and the relevant cost functions may not be lower semi-continuous under practically motivated topologies. The existence of lexicographic minimizers in these settings is therefore treated as an open research question rather than an established result. This limitation does not affect the normative architecture of the framework; it affects the conditions under which the revision procedure is guaranteed to terminate.

A.4 Properties of Robust Discrimination

Let:

$$Discrim(R) = \inf_{\pi \in \Pi_R} D(R, \pi).$$

Since $D(R, \pi) \in [0, 1]$, and $\Pi_R \subseteq Prob(\Theta_R)$, we assume:

- Π_R is weak*-compact,
- $D(R, \pi)$ is continuous in π .

Then the infimum is attained.

This avoids pathological dependence on limiting operating measures.

Compactness is invoked solely to guarantee attainment of the infimum; alternative regularity assumptions would suffice.

A.5 Partition Structure and Refinement

Let \sim_R be defined as in subsection 3.1.

Then \sim_R is an equivalence relation.

Resolution refinement defines a partial order on regimes via partition refinement.

Transitivity of refinement guarantees:

If $R \prec R'$ and $R' \prec R''$ via resolution improvement, then resolution is monotone along chains.

A.6 Invariance Under Reparametrisation

Let $\psi : \mathcal{Y}_R \rightarrow \mathcal{Y}_{R'}$ be a bijection.

Assume:

K_R transforms via pushforward: $K_{R^\psi}(A'|\Theta) = K_R(\psi^{-1}(A')|\Theta)$.

Π_R transforms via induced pushforward.

$Agree_R(p, y) \leftrightarrow Agree_{R^\psi}(\psi_{*p}, \psi(y))$.

Then:

- $EA(T, R) \leftrightarrow EA(T, R^\psi)$,
- $Discrim(R) = Discrim(R^\psi)$,
- Resolution partitions are isomorphic,
- Dom^* is invariant.

Thus, superiority cannot arise from mere reparametrisation.

A.7 Monotonicity of Integrative Burden

For any $T \subseteq T'$ such that T' adds structure inert relative to R :

$$Pred_R(T') = Pred_R(T)$$

,

we assume:

$$pen_R(T') > pen_R(T)$$

.

This ensures strict penalisation of empirically invisible extension.

A.8 Partial Order Properties of Regime Upgrade

Under Dom^* :

- Irreflexivity holds (no strict self-upgrade).
- Transitivity holds if continuity parameters are compatible.
- Antisymmetry holds modulo structural equivalence.

Thus, \prec induces a partial order on equivalence classes of regimes.

A.9 Non-Pathological Regimes

A regime that rejects all models or trivially accepts all models would maximise discrimination vacuously or collapse it entirely.

To prevent such pathologies, we impose a minimal non-triviality constraint:

There exists $\delta > 0$ such that:

$$\forall R \exists T \in \mathcal{T}_{adm} :$$

with

$$\inf_{\pi \in \Pi_R} \mu_{R,\pi}(\{y | \exists m \in Mod(T), g \in \mathcal{G}(T, R) : Agree_R(g(m), y) = 1\}) \geq \delta > 0$$

This ensures that at least one theory remains meaningfully testable under admissible operating conditions.

A reference regime must allow for empirical engagement.

References

- Akaike, Hirotugu (1974). “A new look at the statistical model identification”. In: *IEEE Transactions on Automatic Control* 19, pp. 716–723.
- Balzer, Wolfgang, C. Ulises Moulines, and Joseph D. Sneed (1987). *An Architectonic for Science: The Structuralist Program*. Vol. 186. Synthese Library. Dordrecht: D. Reidel Publishing Company / Springer. DOI: 10.1007/978-94-009-3765-9.
- Chang, Hasok (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press.
- (2011). “The Persistence of Epistemic Objects through Scientific Change”. In: *Erkenntnis* 75, pp. 413–429.
- Duhem, Pierre (1954). *The Aim and Structure of Physical Theory*. Translated by Philip P. Wiener (original French edition 1906). Princeton: Princeton University Press. ISBN: 069102524X.

- Fraassen, Bas C. van (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford: Oxford University Press. ISBN: 0199278229.
- Goldman, Alvin I. (1999). *Knowledge in a Social World*. Oxford: Oxford University Press. ISBN: 0198238207.
- Hacking, Ian (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press. ISBN: 0521282462. DOI: 10.1017/CB09780511814563.
- Kuhn, Thomas S. (1970). *The Structure of Scientific Revolutions*. 2nd ed. Chicago: University of Chicago Press. ISBN: 0226458040.
- Machamer, Peter, Lindley Darden, and Carl F. Craver (2000). “Thinking about Mechanisms”. In: *Philosophy of Science* 67.1, pp. 1–25. DOI: 10.1086/392759.
- Morgan, Mary S. and Margaret Morrison, eds. (1999). *Models as Mediators: Perspectives on Natural and Social Science*. Vol. 52. Ideas in Context. Cambridge: Cambridge University Press. ISBN: 0521650976.
- Nagel, Thomas (1986). *The View from Nowhere*. New York and Oxford: Oxford University Press. ISBN: 0195036689.
- Popper, Karl R. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. Often cited from later revised editions; use the edition you consulted. London: Routledge and Kegan Paul.
- Quine, W. V. O. and J. S. Ullian (1970). *The Web of Belief*. New York: Random House. ISBN: 039430232X.
- Schwarz, Gideon (1978). “Estimating the dimension of a model”. In: *Annals of Statistics* 6, pp. 461–464.
- Stegmüller, Wolfgang (1979). *The Structuralist View of Theories: A Possible Analogue of the Bourbaki Programme in Physical Science*. Berlin and Heidelberg: Springer. DOI: 10.1007/978-3-642-95360-6.
- Suppe, Frederick, ed. (1977). *The Structure of Scientific Theories*. 2nd ed. Urbana and Chicago: University of Illinois Press. ISBN: 0252006348.
- Tal, Eran (2013). “Old and New Problems in Philosophy of Measurement”. In: *Philosophy Compass* 8.12, pp. 1159–1173.
- (2017). *Measurement in Science*. Stanford Encyclopedia of Philosophy. Winter 2017 Edition. URL: <https://plato.stanford.edu/entries/measurement-science/>.