The Pre-Integrative Rejection Principle: A Universal Integrity Filter in Self-Maintaining Systems

Boris Kriger

Institute of Integrative and Interdisciplinary Research

boriskriger@interdisciplinary-institute.org

Abstract

This paper proposes and defends a general systems principle describing a rapid rejection mechanism present in all self-maintaining systems. This mechanism operates prior to full analytical integration and serves to preserve structural integrity by excluding inputs that cannot be safely assimilated. Building on foundational work in autopoiesis theory, cybernetics, and immunology, I argue that this principle represents a necessary architectural feature of viable systems rather than a domain-specific adaptation. The principle is examined across biological organisms, immune systems, ecosystems, computational architectures, cultures, and cognitive frameworks. I formalize the mechanism, propose operational criteria for its identification, and derive testable predictions. System stability is shown to depend on the calibrated threshold of this rejection mechanism: excessively low thresholds produce rigidity and stagnation; excessively high thresholds permit destructive integration and systemic degradation. The paper concludes by examining implications for adaptability, resilience, and long-term sustainability.

Keywords: system integrity, rejection filter, pre-integrative response, systemic stability, adaptive thresholds, structural preservation, autopoiesis, cybernetics, signal detection theory, dual-process cognition

1 Introduction

Self-maintaining systems continuously interact with environments containing both compatible and incompatible elements. The challenge every such system faces is fundamentally one of discrimination: which inputs should be integrated, and which should be excluded? Full analytical evaluation of every input is energetically costly and temporally inefficient. A bacterium that carefully analyzed every molecule in its environment before responding would be outcompeted by one that reacted rapidly to obvious threats. A society that deliberated exhaustively over every foreign practice before permitting or prohibiting it would be paralyzed.

This observation suggests that viable systems require a rapid, low-cost mechanism that excludes potentially destructive inputs before deeper processing occurs. This paper argues that such a mechanism is not merely common but necessary—a universal structural feature of any system capable of maintaining itself over time.

The claim is not trivial. One might object that "systems which survive are those that avoid destruction" is a tautology. But the Pre-Integrative Rejection Principle makes a stronger and more specific claim: that this avoidance takes a particular architectural form, namely a fast filter that operates *before* and *independently of* full analytical processing. This architecture has identifiable properties, generates predictions, and can fail in characteristic ways.

1.1 Relation to Existing Theoretical Frameworks

The principle proposed here builds upon and synthesizes several established traditions in systems theory. I first review the four foundational streams most directly related to PIRP, then survey additional conceptual neighbors, and finally clarify what PIRP adds to this existing landscape.

1.1.1 Core Supporting Traditions

Autopoiesis. Maturana and Varela (Maturana and Varela, 1980) defined autopoietic systems as those that continuously produce and maintain their own boundaries. Autopoietic systems are operationally closed yet structurally open—they continuously produce their own components while rejecting perturbations that would destroy organization. The key insight is that such systems distinguish self from non-self via structural coupling and perturbation rather than representational mirroring. Rejection is not a separate module but intrinsic to the self-production network.

The closest precursor to PIRP in autopoiesis theory is the idea that autopoietic unity requires ongoing discrimination of what can be assimilated without disintegration—cellular membrane selectivity being operational closure in action. PIRP advances this by proposing an explicit two-stage architecture (fast pattern-based rejection → slow analysis), a formal threshold parameter, and cross-domain structural equivalence.

Cybernetics and Ashby's Law. Ashby's Law of Requisite Variety (Ashby, 1956) states that a system's regulatory capacity must match the variety of disturbances it faces. This demands that regulators reduce environmental variety before full processing—exactly the role of a pre-filter. Beer's (Beer, 1972) Viable System Model emphasizes attenuation and filtering at early recursions (System 1–3 interactions) to prevent overload. The Good Regulator theorem (Conant and Ashby, 1970) further establishes that effective control requires modeling, which in turn requires filtering irrelevant variation.

PIRP advances this tradition by making the pre-filter explicitly conservative/biased (false negatives preferred over false positives) and by linking threshold miscalibration to specific failure modes (rigidity versus brittleness).

Immunological Self/Non-Self Distinction. Burnet's (Burnet, 1959) clonal selection theory and subsequent work on immunological tolerance established that the immune system operates through rapid pattern-matching that distinguishes self-markers from foreign antigens. Janeway and Medzhitov (Janeway and Medzhitov, 2002) characterized innate immunity as the paradigmatic case: pattern recognition receptors (PRRs) detect pathogen-associated molecular patterns (PAMPs) in seconds to minutes via surface signatures, triggering rejection/inflammation before adaptive (analytical) immunity engages over days. Autoimmunity represents θ too low (rejecting self); immunodeficiency represents θ too high (accepting pathogens).

PIRP advances this by treating innate immunity as one instantiation of a universal principle rather than a domain-specific biological solution.

Signal Detection Theory. Green and Swets (Green and Swets, 1966) formalized the tradeoffs involved in discriminating signals from noise under uncertainty. The threshold parameter θ in PIRP is formally identical to the decision criterion in SDT; the conservative bias reflects unequal costs of Type I versus Type II errors in integrity-threatening contexts. PIRP applies SDT logic beyond perception to system-level integrity across non-cognitive domains.

1.1.2 Additional Conceptual Neighbors

Several lines of work describe rapid, pre-deep-processing rejection mechanisms, though none generalize as broadly or formally as PIRP:

Innate-like rapid defenses across biology. Beyond mammalian innate immunity, analogous mechanisms appear throughout life: plant immunity (effector-triggered immunity via resistance proteins), invertebrate innate systems (Toll pathways in *Drosophila*), and even CRISPR spacer acquisition in prokaryotes exhibit fast pattern-based rejection before full metabolic or genetic integration.

Predictive processing and active inference. The predictive processing framework (Friston, 2010; Clark, 2013) describes cognition as hierarchical prediction-error minimization. Incoming signals that violate predictions generate "prediction errors" that must either update internal models or be suppressed. Crucially, "precision weighting" determines how much influence prediction errors have on model updating—low-precision errors are effectively rejected before they can modify beliefs. This mechanism is formally similar to the threshold parameter θ in PIRP: precision weighting determines what gets integrated versus what gets filtered out.

The relationship between PIRP and predictive processing deserves careful attention. In a companion paper, I have argued that predictive processing is not merely one computational strategy among many, but the only scalable and evolutionarily stable architecture for systems operating under realistic physical constraints (Kriger, 2026a). The **Predictive Viability Law** formalizes this: for any adaptive system with latency $\tau > 0$ in a punishing, non-stationary environment, continued persistence requires $I(X; Y_{t+\tau}|Y_t) > 0$—internal states must encode information about future environmental states beyond what is available in present observation.

PIRP and the Predictive Viability Law are complementary rather than competing principles. The Predictive Viability Law establishes that systems must maintain predictive information to persist; PIRP establishes that systems must filter inputs before integration to maintain coherence. Together, they describe two necessary conditions for viability:

1. **Predictive requirement:** Internal states must encode information about future states (Predictive Viability Law)

2. **Filtering requirement:** Incoming inputs must be filtered before integration (PIRP)

A system that predicts but does not filter will be overwhelmed by incompatible inputs. A system that filters but does not predict will respond to the past rather than the future. Both requirements must be satisfied for long-term viability in dynamic, punishing environments.

This synthesis suggests a unified architecture: the pre-integrative rejection filter $R(x)$ serves as a gating mechanism that determines which inputs are permitted to update the predictive model. High-precision prediction errors (those signaling genuine model inadequacy) pass the filter; low-precision errors (noise, anomalies, incompatible inputs) are rejected. The threshold θ in PIRP corresponds to precision weighting in predictive processing—both determine the boundary between what gets integrated and what gets filtered out.

Cognitive science dual-process models. Kahneman's (Kahneman, 2011) System 1 / System 2 distinction describes rapid, automatic cognitive processes that can reject incoherence before slower deliberation engages. Haidt's (Haidt, 2001) social intuitionist model treats moral intuitions as automatic vetoes that precede and often override moral reasoning. Phenomena such as motivated reasoning and belief perseverance function as cognitive integrity filters. Crucially, these are usually treated in the literature as biases to overcome rather than architectural necessities—PIRP's reframing as structural features is distinctive.

A note of caution is warranted here. The psychological literature on dual-process cognition is contested. Critics of Haidt's moral intuitionism argue that reasoning plays a larger role in moral judgment than the "emotional dog and its rational tail" model suggests (Paxton and Greene, 2010). More fundamentally, there is ongoing debate about whether "System 1" and "System 2" represent distinct cognitive systems or merely different modes of processing within a unified system (Melnikoff and Bargh, 2018). PIRP does not depend on resolving these debates—it requires only that cognitive systems exhibit *some* rapid filtering mechanism that satisfies the operational criteria, not that this mechanism corresponds to a neurologically distinct "system."

Purity, pollution, and disgust. Mary Douglas's (Douglas, 1966) anthropological work on purity and pollution frameworks describes how cultures rapidly categorize and reject anomalous phenomena that threaten classificatory systems. Rozin and Fallon's (Rozin and Fallon, 1987) research on disgust reveals it as a mechanism for rapid rejection of potential contaminants—both biological and moral—before deliberative evaluation.

Computational security and input validation. Firewalls, intrusion detection systems, and anomaly-based filters reject network traffic based on headers and signatures before payload analysis. Input sanitization in software rejects malformed data before processing. These represent almost literal implementations of the Stage 1 $\rightarrow$ Stage 2 model proposed in PIRP.

Ecosystem invasion biology. Biotic resistance via competitive exclusion, predation, parasitism, and allelopathy acts as distributed, pre-adaptive rejection of invasive species. This represents the closest match to PIRP's treatment of distributed systems without central controllers.

Chemical Organization Theory. Recent work in the tradition of Fontana and Buss (Fontana and Buss, 1994) on algorithmic chemistry, extended by Dittrich and colleagues into Chemical Organization Theory (Dittrich et al., 2001; Kreyssig et al., 2012), examines how self-maintaining reaction networks exhibit closure and exclude disruptive molecular species—a formal-chemical analog to rejection before integration.

Epistemic gatekeeping. Scientific communities maintain integrity through mechanisms such as peer review, editorial filtering, and replication requirements that operate before claims are integrated into accepted knowledge. These mechanisms exhibit PIRP-like properties: they are faster and cheaper than full verification, they operate on surface features (methodology, statistical reporting, author credentials), they are biased toward rejection (most submissions are rejected), and they can be overridden by higher-order processes (editorial discretion, appeals).

Institutional exclusion and non-participation. In a parallel formal treatment, I have shown that institutional stability may be better preserved through early exclusion and non-participation than through attempted accommodation of incompatible external engagements (Kriger, 2026b). The formal framework defines net institutional value as $V(c) = R(c) - C(c)$, where $C(c)$ includes interaction costs, coordination costs, and critically, *irreversible institutional distortion* $D(c)$. The key result—that early exclusion dominates delayed exit for engagements with $V(c) < 0$—is structurally identical to PIRP's claim that pre-integrative rejection is more efficient than post-integration correction.

Organizational learning and defensive routines. Argyris and Schön's (Argyris and Schön, 1978) work on organizational learning describes "defensive routines"—organizational patterns that prevent the detection and correction of errors. These routines function as rejection mechanisms that protect organizational identity by excluding threatening information.

1.1.3 What PIRP Adds

The existing literature is rich in domain-specific mechanisms but lacks several features that PIRP provides:

1. **A single, abstract, falsifiable principle** that applies equally to cells, immune systems, ecosystems, firewalls, cultures, and minds—not merely analogical comparison but claimed structural equivalence.

2. **Explicit operational criteria** (temporal, resource, independence, signature, override) for recognizing the mechanism in any domain.

3. **A unified treatment of threshold miscalibration pathologies** across domains, with the same diagnostic framework applicable to autoimmune disease, organizational rigidity, and cultural fundamentalism.

4. **The fractal/nested scaling argument** with cross-level interactions and override costs (e.g., professional suppression of disgust responses).

5. **The strong claim of structural equivalence**—not homology (shared evolutionary origin) or loose analogy, but identical abstract information-processing architecture realized through different physical substrates.

PIRP is thus best characterized as original theoretical synthesis rather than rediscovery. It takes well-established pieces (autopoietic closure, variety attenuation, innate immunity, SDT thresholds, moral intuitions, firewall logic) and assembles them into a novel, generalizable principle with new predictive and diagnostic power.

2 Definition of the Pre-Integrative Rejection Filter

Terminological note: I use "pre-integrative" rather than "pre-analytical" throughout this paper, following reviewer feedback that "pre-analytical" is misleading since the rejection function $R(x)$ itself performs a kind of analysis (pattern-matching). "Pre-integrative" more accurately captures the mechanism's role: filtering before integration, not before any analysis whatsoever. The function $R(x)$ analyzes surface features; what it precedes is the deeper analytical processing that would be required for integration.

2.1 Formal Specification

Let S be a self-maintaining system interacting with environment E. For an incoming element $x \in E$, the system must decide whether integration is permissible. I propose that this decision occurs in two stages:

Stage 1: Pre-Integrative Rejection. Before full analytical evaluation $A(x)$, the system applies a rapid classification function:

$$R(x) = \begin{cases} 0, & \text{reject immediately} \\ 1, & \text{allow further processing} \end{cases} \tag{1}$$

Stage 2: Analytical Evaluation. For elements where $R(x) = 1$, the system applies a more costly analytical function $A(x)$ that determines the mode and degree of integration.

The pre-integrative function $R(x)$ has the following properties:

1. **Speed:** $R(x)$ completes faster than $A(x)$
2. **Cost:** $R(x)$ requires fewer resources than $A(x)$
3. **Conservatism:** $R(x)$ is biased toward rejection (false negatives are less costly than false positives for system integrity)
4. **Pattern-based:** $R(x)$ operates on surface features or learned signatures rather than deep structural analysis

2.2 The Threshold Parameter

The rejection function depends on a threshold parameter θ that determines the boundary between rejection and acceptance:

$$R(x) = \begin{cases} 0, & \text{if } \phi(x) < \theta \\ 1, & \text{if } \phi(x) \geq \theta \end{cases} \tag{2}$$

Here $\phi(x)$ represents a compatibility score—a rapid assessment of how closely x matches patterns associated with safe integration. The threshold θ determines how high this score must be for x to pass to analytical evaluation.

2.3 Operational Criteria

For the principle to be non-trivial and testable, we need criteria for identifying pre-integrative rejection in a given system. I propose the following:

1. **Temporal criterion:** The rejection response occurs before full processing would complete
2. **Resource criterion:** The rejection response consumes fewer resources than full processing
3. **Functional independence criterion:** The rejection response can occur even when full analytical capacity is available (this refers to *functional* independence—the rejection logic is separate from analytical logic—not necessarily *resource* independence)

4. **Signature criterion:** The rejection response is triggered by identifiable surface features rather than deep structural properties

5. **Override criterion:** The rejection response can be suppressed by higher-order control, but this suppression is effortful

A mechanism satisfies the Pre-Integrative Rejection Principle if it meets all five criteria.

2.4 Defining Self-Maintenance and System Identity

To avoid the charge of tautology—that PIRP merely restates "systems that survive are those that avoid what would destroy them"—we need a more rigorous definition of the "self" in self-maintenance.

Self-maintenance refers to the continuous process by which a system preserves the organization of relationships among its components, even as components themselves may change. This definition, drawn from autopoiesis theory, distinguishes between:

- **Identity:** The specific components that constitute the system at a given time

- **Organization:** The pattern of relationships among components that defines the system's type

A self-maintaining system preserves its *organization* while potentially replacing its *identity* (specific components). A cell maintains its organization as a living cell while continuously replacing its molecular components. A culture maintains its organization as a distinct culture while its individual members change.

This clarification also addresses why PIRP is not trivially true. The principle does not merely claim that surviving systems avoid destruction. It claims that self-maintenance requires a *specific architectural feature*—a fast, pattern-based filter operating before analysis—and that the calibration of this filter determines system stability. A system might theoretically maintain itself through full analytical processing of all inputs; PIRP claims such systems are not viable because the costs of full analysis exceed available resources. This is an empirical claim that could be false.

3 Cross-Domain Manifestations

This section examines how the pre-integrative rejection mechanism manifests across different system types. The claim is not merely that rejection occurs in all these domains—that would be trivial—but that rejection in each domain exhibits the five operational criteria specified above.

3.1 Biological Systems

Cellular level. Cell membranes employ receptor-mediated selectivity that rejects molecules lacking appropriate binding signatures before any metabolic processing occurs. This rejection is fast (milliseconds), low-cost (no ATP expenditure for rejection itself), pattern-based (depends on molecular shape), and can be modulated by cellular state (criterion 5).

Organismic level. Vomiting, coughing, and sneezing reflexes expel potentially harmful substances before absorption or deep infection. These responses are triggered by surface features (bitter taste, irritation, foreign particles) rather than by analysis of actual toxicity. The speed and automaticity of these responses, combined with their susceptibility to voluntary suppression (one can sometimes inhibit a cough), satisfies all five criteria.

Toxicological evidence. Organisms routinely reject substances that would be harmless or even beneficial if analyzed. Bitter-taste aversion causes rejection of many non-toxic plants. This "false positive" pattern is predicted by the principle: a conservative pre-integrative filter will reject some compatible inputs to ensure rejection of incompatible ones.

3.2 Immune Systems

The immune system provides the clearest biological paradigm of pre-integrative rejection. Innate immunity responds within minutes to pathogen-associated molecular patterns (PAMPs) without requiring the slower adaptive immune response. Pattern recognition receptors (PRRs) detect conserved microbial signatures and trigger immediate inflammatory and phagocytic responses.

This mechanism satisfies all five criteria:

- **Temporal:** Innate responses precede adaptive responses by days
- **Resource:** Pattern recognition requires less cellular machinery than antibody production
- **Independence:** Innate immunity operates even when adaptive immunity is impaired
- **Signature:** PRRs recognize surface patterns (lipopolysaccharides, flagellin) not functional properties
- **Override:** Regulatory T cells and anti-inflammatory signals can suppress innate responses

The existence of autoimmune diseases demonstrates what happens when the threshold θ is miscalibrated: the system rejects elements (self-tissues) that should be accepted.

3.3 Ecosystems

Ecosystems exhibit rejection of invasive species through multiple mechanisms: competitive exclusion, predation, parasitism, and allelopathy (chemical inhibition). These responses are not "planned" but emerge from the existing network of interactions.

The ecosystem case raises an important clarification: the "system" performing rejection is distributed rather than unified. There is no central controller deciding to reject an invasive species. Rather, rejection emerges from the aggregate behavior of existing species whose interactions are disrupted by the invader.

This distributed rejection still satisfies the operational criteria:

- **Temporal:** Initial resistance (predation, competition) precedes long-term adaptation

- **Resource:** Existing interactions require no new resource allocation
- **Independence:** Resistance occurs regardless of ecosystem "analysis" of the invader
- **Signature:** Responses are triggered by the invader's observable behaviors and resource use
- **Override:** Human intervention can suppress natural resistance (e.g., removing predators)

Failed rejection—successful invasion—occurs when the invasive species falls outside the recognition patterns of existing rejection mechanisms. This is precisely what the principle predicts: rejection filters tuned to familiar threats may fail against novel ones.

3.4 Computational Systems

Network security provides clear examples of pre-integrative rejection:

Firewalls reject packets based on header information (source IP, port numbers) before any analysis of payload content. A packet from a blacklisted IP is dropped in microseconds; analyzing its contents would take orders of magnitude longer.

Input validation rejects malformed data before processing. A SQL query containing suspicious characters is rejected before the database evaluates whether the query would actually cause harm.

Rate limiting rejects requests that exceed frequency thresholds regardless of their content.

These mechanisms satisfy all five criteria. Notably, they can be disabled by administrators (override criterion), but this is recognized as a security risk—demonstrating that the filter serves a protective function.

The computational domain also illustrates threshold miscalibration. Overly aggressive spam filters (low θ) reject legitimate emails. Overly permissive filters (high θ) allow malware. The ongoing arms race between attackers and defenders is fundamentally about shifting the boundary between rejection and acceptance.

3.5 Cultural Systems

Cultures exhibit rejection of norm-violating behaviors through social sanctions: disgust, ostracism, ridicule, and punishment. These responses are often rapid, automatic, and pattern-based.

Mary Douglas's (Douglas, 1966) seminal anthropological work on purity and pollution illuminates this mechanism. Douglas demonstrated that cultures rapidly categorize and reject phenomena that threaten classificatory systems—"matter out of place" that violates cultural categories triggers immediate rejection responses before any deliberative evaluation of actual harm. This is structurally identical to the pattern-based rejection described in PIRP.

Rozin and Fallon's (Rozin and Fallon, 1987) research on disgust further supports this analysis. Disgust functions as a rapid rejection mechanism for potential contaminants—initially biological (spoiled food, bodily fluids) but extended through cultural evolution to moral violations. The "core disgust" response operates in milliseconds based on surface features (appearance, smell, texture) before any analysis of actual toxicity or harm.

Nevertheless, the operational criteria can be adapted:

- **Temporal:** Disgust and disapproval are immediate; deliberative moral judgment takes longer
- **Resource:** Automatic social sanctions require less cognitive effort than ethical analysis
- **Independence:** People experience moral intuitions even when they cannot articulate reasons
- **Signature:** Initial reactions are triggered by surface features of behavior, not deep analysis of consequences
- **Override:** Cultural training and deliberate reflection can override initial reactions

Haidt's (Haidt, 2001) work on moral intuitions supports this analysis. His research shows that moral judgments are often made rapidly and automatically, with reasoning following as post-hoc justification. This is precisely the architecture predicted by the Pre-Integrative Rejection Principle.

3.5.1 Case Study: Implicit Bias as Pre-Integrative Rejection

Research on implicit bias provides a particularly well-documented example of cultural $R(x)$ mechanisms and their potential miscalibration. The Implicit Association Test (IAT) and related measures demonstrate that evaluative responses to social categories (race, gender, age) occur within 200–300 milliseconds—far before conscious, deliberative processing ($A(x)$) can intervene (Greenwald et al., 1998).

These rapid evaluations satisfy all five operational criteria:

- **Temporal:** IAT responses occur in ∼300 ms; deliberative responses take ∼2,000+ ms
- **Resource:** Implicit processing requires less cognitive effort than explicit reasoning
- **Functional independence:** Implicit and explicit attitudes can diverge, indicating separate processing
- **Signature:** Responses are triggered by surface features (phenotype, accent, dress)
- **Override:** With effort and motivation, implicit responses can be modulated (though not eliminated)

From the PIRP perspective, implicit biases represent the operation of culturally calibrated rejection thresholds. In societies with histories of intergroup conflict or segregation, $R(x)$ has been calibrated to lower thresholds for out-group members—a configuration that may have had adaptive value in ancestral environments but produces pathological outcomes in diverse modern societies.

This analysis suggests that implicit bias is not merely a cognitive error to be corrected but a manifestation of threshold miscalibration requiring recalibration. The distinction matters for intervention design: error correction implies teaching people the "right" answer, while recalibration implies providing experiences that shift the underlying threshold parameters. The latter approach aligns with evidence that sustained intergroup contact is more effective than explicit instruction in reducing implicit bias (Pettigrew and Tropp, 2006).

3.5.2 Societal Stress and Threshold Dynamics

Cross-cultural research supports the prediction that environmental stress affects rejection thresholds. Societies experiencing high threat levels (war, economic instability, pandemic) show measurable decreases in θ—increased rejection of foreign inputs, out-groups, and novel practices (Gelfand et al., 2011).

This pattern is precisely what PIRP predicts. Under resource scarcity or elevated threat, the costs of false positives (accepting harmful inputs) increase relative to the costs of false negatives (rejecting beneficial ones). Rational threshold adjustment lowers θ, increasing conservatism. While adaptive in the short term, sustained low θ can produce rigidity and xenophobia that impair long-term adaptation.

3.6 Cognitive Systems

Individual cognition exhibits pre-integrative rejection of incoherent or threatening information. This manifests as:

- Immediate dismissal of claims that contradict core beliefs
- Difficulty processing statements that violate semantic expectations
- Resistance to information that threatens self-concept

Cognitive science provides extensive documentation of these phenomena under labels such as confirmation bias, motivated reasoning, and belief perseverance. What the Pre-Integrative Rejection Principle adds is a functional interpretation: these "biases" are not mere errors but implementations of a necessary integrity-preservation mechanism.

This interpretation is supported by research on cognitive load. When analytical resources are depleted, people rely more heavily on heuristic rejection. This is predicted by the principle: the pre-integrative filter should operate even when full analysis is unavailable.

The override criterion is also satisfied: deliberate effort can overcome initial rejection of dissonant information. Education and training can recalibrate the threshold, making previously rejected ideas acceptable for consideration. But the filter itself is never eliminated—only its parameters change.

4 Structural Equivalence: A Precise Claim

Having surveyed manifestations across domains, I now address directly the claim that these processes are "structurally equivalent." This claim requires precision and some qualification.

4.1 What Structural Equivalence Means (and Does Not Mean)

I do not claim that immune rejection and cultural rejection involve the same physical mechanisms. Obviously they do not. Nor do I claim that insights from one domain can be directly transferred to another without careful translation.

The claim is rather that both instantiate the same *abstract functional architecture*:

1. A compatibility-assessment function ϕ that rapidly evaluates inputs against stored patterns
2. A threshold parameter θ that determines the acceptance boundary
3. A binary rejection decision R that occurs before full analytical processing
4. Characteristic failure modes when θ is miscalibrated

This is the sense in which the processes are structurally equivalent: they share an abstract information-processing architecture even though they differ in physical implementation.

Important qualification: This claim of structural equivalence is stronger than mere analogy but weaker than identity. Analogy notes surface similarities; identity requires sameness at all levels; structural equivalence claims sameness of abstract architecture with acknowledged differences in implementation. Whether this intermediate claim is useful depends on whether the shared architecture generates predictions and insights that domain-specific analyses miss.

4.2 Criteria for Structural Equivalence

To avoid the charge that the mapping is merely stipulative, I propose the following criteria for structural equivalence:

1. **Formal correspondence:** Components can be mapped between domains such that the functional relationships among components are preserved
2. **Parallel predictions:** The abstract architecture generates the same predictions in both domains (e.g., conservative bias, threshold-dependent stability)
3. **Parallel pathologies:** Miscalibration produces analogous failure modes in both domains
4. **Independent verification:** The architecture can be identified in each domain without prior knowledge of the other

The examples in Section 3 are intended to demonstrate that these criteria are met. Immune systems and cognitive systems independently exhibit fast pattern-based filtering, threshold-dependent behavior, conservative bias, and characteristic pathologies of miscalibration. The question is whether this convergence reflects a genuine underlying principle or mere coincidence.

4.3 Mapping Across Domains

The following table makes the structural equivalence precise by identifying the components of the pre-integrative rejection architecture in each domain:

5 Threshold Calibration and System Stability

The stability and adaptability of a system depend critically on the calibration of its rejection threshold θ.

Domain	System S	Input x	Pattern $\phi(x)$	Threshold θ	Rejection
Cellular	Cell	Molecule	Receptor binding	Membrane selectivity	Non-bindin
Immune	Organism	Pathogen	PAMP recognition	Tolerance threshold	Inflammati
Ecosystem	Community	Species	Niche overlap	Competitive capacity	Exclusion
Computational	Network	Packet	Header matching	Firewall rules	Drop
Cultural	Society	Behavior	Norm matching	Social tolerance	Sanction
Cognitive	Mind	Proposition	Coherence check	Belief threshold	Dismissal

Table 1: Mapping of PIRP components across domains

5.1 Costs of Miscalibration

Low θ (excessive rejection):

- Rigidity: the system cannot adapt to changing conditions
- Missed opportunities: beneficial inputs are rejected
- Isolation: the system becomes closed to its environment
- Brittleness: lacking adaptive capacity, the system may fail catastrophically when rejection proves insufficient

High θ (excessive acceptance):

- Contamination: harmful inputs are integrated
- Loss of coherence: the system's identity becomes diffuse
- Degradation: accumulated incompatible elements impair function
- Dissolution: in extreme cases, the system ceases to exist as a distinct entity

5.2 The Optimal Range

System viability requires θ to remain within an optimal range:

$$\theta_{\min} < \theta < \theta_{\max} \tag{3}$$

The boundaries of this range are determined by:

- The distribution of inputs in the environment (what the system typically encounters)
- The costs of integration failure versus rejection failure
- The system's capacity for recovery from integration errors

This framing connects to signal detection theory. The optimal threshold depends on the base rates of compatible versus incompatible inputs and the relative costs of false positives versus false negatives. In most natural systems, false positives (accepting harmful inputs) are more costly than false negatives (rejecting beneficial ones), which predicts a conservative bias—consistent with observed behavior.

5.3 Quantifying Resource Asymmetry

A key empirical prediction of PIRP is that rejection ($R(x)$) consumes fewer resources than full analysis ($A(x)$). The following table provides concrete examples of this asymmetry across domains:

Domain	$R(x)$ **Process**	$A(x)$ **Process**	**Time Ratio**
Network security	Header inspection	Deep packet inspection	$\sim 10^{-6}$ s vs. $\sim 10^{-3}$ s
Immune (innate vs. adaptive)	PRR recognition	Antibody production	Minutes vs. days
Cognitive (perception)	Pre-attentive filtering	Conscious analysis	~100 ms vs. ~500 ms
Cognitive (moral)	Disgust response	Ethical deliberation	~200 ms vs. ~2,000 m
Email filtering	Spam header check	Content analysis + ML	~1 ms vs. ~100 ms
Taste aversion	Bitter receptor activation	Toxicological analysis	~50 ms vs. N/A

Table 2: Resource asymmetry between $R(x)$ and $A(x)$ across domains

These ratios are approximate and context-dependent, but they demonstrate a consistent pattern: pre-integrative rejection is orders of magnitude faster and cheaper than full analytical processing. This asymmetry is what makes the two-stage architecture adaptive—it conserves resources by preventing full analysis of inputs that can be adequately filtered by surface features.

5.4 Dynamic Calibration

In healthy systems, θ is not static but dynamically calibrated based on experience and context. This calibration occurs through multiple mechanisms:

- **Learning:** Exposure to inputs adjusts the patterns used for compatibility assessment
- **Contextual modulation:** Different contexts activate different threshold settings
- **Feedback:** Integration failures and successes adjust future thresholds
- **Developmental change:** Thresholds typically change across the system's lifespan

The immune system provides a clear example. During fetal development and early infancy, the immune system "learns" self-markers and adjusts its tolerance threshold. Failure of this calibration produces autoimmune disease (θ too low, rejecting self) or immunodeficiency (θ too high, accepting pathogens).

6 Fractal Scaling Across Organizational Levels

A striking feature of the pre-integrative rejection mechanism is its appearance at multiple scales of organization. Cells within organisms, organisms within ecosystems, individuals within societies—each level exhibits its own rejection filter, and these filters are nested within each other.

6.1 The Nested Architecture

Consider a human within a society:

- **Cellular level:** Each cell rejects incompatible molecules
- **Immunological level:** The immune system rejects pathogens
- **Cognitive level:** The mind rejects incoherent information
- **Social level:** The community rejects norm-violating behavior

These filters are not independent but interact. Cognitive rejection of certain information may be partly driven by anticipation of social rejection. Immune responses can affect cognitive states. Social structures shape what cognitive thresholds develop.

6.2 Cross-Level Calibration

The nested structure creates both redundancy and complexity. Redundancy provides robustness: if one level's filter fails, others may compensate. A toxin that passes cellular defenses may be expelled by organismic reflexes; a harmful idea that passes cognitive filters may be sanctioned socially.

But the nested structure also creates potential for conflict. Social pressure may force cognitive acceptance of propositions that would otherwise be rejected. Professional training may require overriding disgust responses that would normally trigger rejection. Organisms may tolerate symbionts that cellular mechanisms would reject.

This cross-level interaction suggests that system health depends not only on the calibration of individual filters but on their coordination across levels.

7 Temporary Suppression and Professional Contexts

An important feature of the pre-integrative rejection mechanism is that it can be temporarily suppressed by higher-order control while remaining architecturally intact. Professional contexts provide clear examples.

7.1 Medical Professionals

Surgeons must suppress disgust responses that would normally trigger rejection of exposure to blood, tissue, and bodily fluids. This suppression is effortful and must be learned. Medical students report initial strong rejection responses that diminish with training. But the underlying mechanism is not eliminated—it is overridden. Evidence for this includes:

- Surgeons still experience disgust in non-professional contexts
- Under extreme stress or fatigue, the suppressed response may reassert itself
- The suppression requires ongoing cognitive effort, even if it becomes automatized

7.2 Forensic Investigators

Investigators examining crime scenes must suppress both disgust (at gruesome scenes) and moral rejection (of the crimes committed). This requires maintaining analytical distance while the pre-integrative rejection system signals "reject, reject, reject."

7.3 Theoretical Implications

The suppressibility of pre-integrative rejection demonstrates several important points:

1. The rejection mechanism is *architectural*, not merely a surface reaction. A surface reaction would simply be eliminated by training; an architectural feature can only be overridden.

2. The override capacity is itself a system feature that can be developed. Systems are not passive recipients of their rejection responses but can modulate them through meta-level control.

3. Chronic override may have costs. Research on "compassion fatigue" and burnout in helping professions suggests that continuous suppression of rejection responses is depleting and can produce pathological outcomes.

4. The existence of override mechanisms suggests that pre-integrative rejection is *calibrated for typical environments*. Professional contexts represent atypical environments where the default calibration is inappropriate and must be adjusted.

8 Adaptation as Threshold Reconfiguration

Long-term adaptation—whether evolutionary, developmental, or cultural—does not eliminate the pre-integrative rejection mechanism but reconfigures its parameters. This section examines how adaptation occurs while preserving the underlying architecture.

8.1 Evolutionary Adaptation

Natural selection adjusts rejection thresholds across generations. Populations exposed to specific pathogens evolve immune systems tuned to those threats. Populations in novel environments may exhibit maladapted rejection responses that require generations to recalibrate.

Example: The evolution of lactose tolerance in human populations with dairying traditions represents a recalibration of digestive rejection mechanisms to accept inputs that would trigger rejection in other populations.

8.2 Developmental Adaptation

Individual organisms adjust rejection thresholds through development. This is particularly clear in:

- **Immune maturation:** The developing immune system learns self-markers and adjusts tolerance

- **Taste preferences:** Children reject bitter tastes more strongly than adults; calibration shifts with exposure
- **Cognitive development:** Children's cognitive filters become more sophisticated with experience and education

8.3 Cultural Adaptation

Cultures adapt their rejection thresholds through collective processes:

- **Generational change:** Each generation's threshold reflects its formative experiences
- **Institution building:** Formal institutions embody and stabilize particular threshold settings
- **Ideological shift:** Cultural movements can rapidly recalibrate what is acceptable for consideration

The civil rights movement, for example, can be understood as a collective recalibration of cultural rejection thresholds—making previously rejected claims (about racial equality) acceptable for serious consideration.

8.4 The Persistence of Architecture

Crucially, all these forms of adaptation preserve the underlying rejection architecture. What changes is the threshold θ and the patterns used for compatibility assessment $\phi(x)$. The binary rejection function $R(x)$ and its priority over analytical processing remain constant.

This observation has implications for understanding both change and resistance to change. Change is possible because thresholds are adjustable. Change is difficult because the architecture creates a default bias toward rejection of unfamiliar inputs. Successful adaptation requires somehow making the unfamiliar familiar enough to pass the initial filter while the analytical system evaluates its actual merits.

9 Implications for Systemic Degradation and Resilience

The Pre-Integrative Rejection Principle has implications for understanding how systems degrade and how they maintain resilience.

9.1 Pathways to Degradation

Systems may degrade through several pathways related to rejection threshold miscalibration:

Threshold collapse ($\theta \to \infty$): The system loses the ability to reject anything. All inputs are passed to analytical processing, which becomes overwhelmed. Harmful elements are integrated, producing cumulative damage. This pathway characterizes certain immune deficiencies, organizational collapse under information overload, and cultural dissolution.

Threshold rigidity ($\theta \to 0$): The system rejects virtually everything. It becomes isolated from environmental feedback, unable to adapt. This pathway characterizes autoimmune diseases, organizational paranoia, and cultural fundamentalism.

Threshold oscillation: The system alternates between excessive rejection and excessive acceptance, never establishing stable calibration. This produces chaotic dynamics and unpredictable behavior.

Threshold mismatch: The system's threshold is calibrated to an environment different from the one it currently inhabits. Inputs that should be rejected are accepted; inputs that should be accepted are rejected. This is common after rapid environmental change.

9.2 Resilience Mechanisms

Resilient systems maintain rejection threshold integrity through several mechanisms:

- **Redundancy:** Multiple nested rejection filters provide backup when one fails
- **Feedback:** Integration outcomes adjust future thresholds
- **Variability:** Heterogeneous threshold settings across subsystems allow some subsystems to accept what others reject, enabling exploration while maintaining overall integrity
- **Meta-regulation:** Higher-order mechanisms monitor and adjust threshold settings

9.3 Predictive Value

The principle generates predictions about system vulnerability:

1. Systems that have recently experienced rapid environmental change are at elevated risk of threshold mismatch
2. Systems with homogeneous threshold settings are more vulnerable to novel threats than heterogeneous systems
3. Systems that suppress feedback about integration outcomes will drift toward miscalibration
4. Systems under resource stress will rely more heavily on pre-integrative rejection, becoming more conservative

These predictions are testable across domains and suggest interventions for improving system resilience.

10 Testable Predictions and Falsifiability

A principle that cannot be tested is not a scientific contribution. This section specifies predictions that follow from the Pre-Integrative Rejection Principle and conditions that would falsify it.

10.1 Positive Predictions

The principle predicts:

1. **Temporal ordering:** In any self-maintaining system, rejection of incompatible inputs will occur faster than full analytical processing. This is testable by measuring response latencies.
   - *Quantitative prediction*: $R(x)$ latency should be at least 2× faster than $A(x)$ latency in any domain; in most domains, the ratio should exceed 10×.
2. **Resource asymmetry:** Rejection will require fewer resources than analysis. This is testable by measuring metabolic costs, computational cycles, or cognitive load.
   - *Quantitative prediction*: $R(x)$ should consume less than 50% of the resources required by $A(x)$; in most domains, less than 10%.
3. **Conservative bias:** Systems will exhibit more false negatives than false positives relative to an optimal decision rule.
   - *Quantitative prediction*: The ratio of false negatives to false positives should exceed 1:1 in integrity-critical contexts; in high-stakes domains (immune, security), should exceed 10:1.
4. **Threshold plasticity:** Rejection thresholds will adjust based on experience while the rejection mechanism itself remains constant.
5. **Override costs:** Suppressing rejection responses will be effortful and depleting.
   - *Quantitative prediction*: Tasks requiring override of $R(x)$ should show elevated markers of cognitive effort compared to tasks where $R(x)$ is aligned with desired behavior.
6. **Calibration pathology:** Systems with miscalibrated thresholds will exhibit characteristic dysfunctions (rigidity for low θ, degradation for high θ).
7. **Stress-induced conservatism:** Under resource scarcity or elevated threat, systems will lower θ (become more rejecting).
   - *Quantitative prediction*: Measurable increase in rejection rates under experimentally induced stress conditions.

10.2 Falsification Conditions

The principle would be falsified by:

1. **A self-maintaining system that lacks any pre-integrative rejection mechanism.** Such a system would analyze all inputs fully before deciding whether to integrate them, with no rapid filter stage.
2. **A self-maintaining system whose rejection mechanism operates more slowly or at greater cost than its analytical mechanism.** This would contradict the core claim about the architecture of rejection.

3. **A system that adapts by eliminating its rejection mechanism rather than recalibrating it.** If adaptation could remove the rejection filter entirely rather than adjust its parameters, the principle would be wrong.

4. **A system where rejection thresholds cannot be miscalibrated, or where miscalibration produces no dysfunction.** The principle claims that threshold calibration matters; if it doesn't, the principle is false.

5. **Absence of conservative bias in a self-maintaining system facing integrity threats.** If a system exhibits equal or greater false positives than false negatives when integration failure is more costly than rejection failure, the SDT-based predictions of PIRP are falsified.

10.3 Addressing the Unfalsifiability Objection

Critics may object that the falsification conditions are impossible to meet in practice—that any system processing inputs selectively could be redescribed as having a "fast filter" at some level. This objection has merit and deserves a direct response.

The objection would be decisive if PIRP claimed merely that "systems reject harmful inputs." But PIRP makes stronger, more specific claims:

1. **Architectural claim:** Rejection occurs via a distinct mechanism ($R(x)$) that is *functionally separable* from analytical processing ($A(x)$), not merely a quick version of the same process.

2. **Temporal claim:** $R(x)$ completes before $A(x)$ begins, not concurrently or after.

3. **Resource claim:** $R(x)$ and $A(x)$ have different resource profiles, with $R(x)$ always cheaper.

4. **Bias claim:** $R(x)$ is conservative, favoring false negatives over false positives.

These claims are empirically testable. I acknowledge that finding counterexamples may be difficult—precisely because PIRP may be correct. But difficulty of falsification is not the same as impossibility. The predictions are specific enough that counterexamples, if they exist, would be recognizable.

11 Formal Statement of the Principle

Based on the foregoing analysis, I offer the following formal statement:

The Pre-Integrative Rejection Principle (PIRP)

Any self-maintaining system S possesses a mechanism R satisfying the following conditions:

1. *R is a function from environmental inputs to $\{accept\ for\ analysis, reject\}$*

2. *R operates prior to and independently of the system's full analytical processing*

3. *R is faster and less resource-intensive than full analytical processing*

4. *R operates on surface features or learned signatures rather than deep structural properties*

5. *R can be suppressed by higher-order control but cannot be eliminated*

6. *The system's stability depends on appropriate calibration of R's threshold parameter*

Systems that maintain well-calibrated rejection thresholds exhibit resilience and adaptive capacity. Systems with miscalibrated thresholds exhibit characteristic pathologies: excessive rejection produces rigidity and isolation; insufficient rejection produces contamination and degradation.

11.1 Relationship to the Predictive Viability Law

PIRP can be formally connected to the Predictive Viability Law (Kriger, 2026a), which states that for any adaptive system with latency $\tau > 0$ in a punishing, non-stationary environment:

$$I(X; Y_{t+\tau}|Y_t) > 0 \tag{4}$$

That is, internal states must encode information about future environmental states beyond what is available in present observation.

Theorem 1 (Joint Viability Theorem). *A self-maintaining system S in a dynamic, punishing environment must satisfy both:*

1. ***Predictive condition:*** $I(X; Y_{t+\tau}|Y_t) > 0$ *(Predictive Viability Law)*

2. ***Filtering condition:*** *There exists $R(x)$ satisfying PIRP conditions 1–6*

Proof sketch. Without prediction (condition 1 violated), the system responds to past states and faces exponentially decaying survival probability. Without filtering (condition 2 violated), the system must fully analyze all inputs. Given finite processing capacity and input arrival rate exceeding processing rate, the system accumulates unprocessed inputs indefinitely, leading to either (a) queue overflow and system failure, or (b) degraded processing quality that violates condition 1. Therefore, both conditions are necessary for persistence. □

Corollary 1 (Filtering-Prediction Trade-off). *For fixed processing capacity C, there exists an optimal allocation between predictive modeling and pre-integrative filtering. Excessive investment in filtering ($\theta \to 0$) starves the predictive model of data needed to maintain $I > 0$. Excessive investment in prediction ($\theta \to \infty$) overwhelms the model with noise and incompatible inputs.*

This formal connection suggests that PIRP and the Predictive Viability Law are two aspects of a single, more fundamental principle governing the architecture of viable systems under physical constraints.

11.2 Relationship to Institutional Exclusion

The formal framework for institutional exclusion (Kriger, 2026b) provides an independent derivation of PIRP-like principles from organizational economics. For an institution facing optional engagement c:

$$V(c) = R(c) - [I(c) + K(c) + D(c)] \tag{5}$$

where $R(c)$ is revenue, $I(c)$ is interaction cost, $K(c)$ is coordination cost, and $D(c)$ is irreversible institutional distortion.

The key result—that early exclusion dominates delayed exit for engagements with $V(c) < 0$—maps directly onto PIRP:

Institutional Framework	**PIRP Framework**
Engagement c	Input x
Revenue $R(c)$	Potential benefit of integration
Distortion cost $D(c)$	Cost of false positive
Net value $V(c)$	Compatibility score $\phi(x) - \theta$
Early exclusion	$R(x) = 0$ (reject)
Theorem: Early exclusion dominates	PIRP: Pre-integrative rejection is optimal

Table 3: Mapping between institutional exclusion and PIRP frameworks

The convergence of these independent analyses—from physical constraints (Predictive Viability Law), from systems architecture (PIRP), and from organizational economics (institutional exclusion)—suggests we are identifying a genuine structural feature of viable systems rather than an artifact of any particular framework.

12 Implications and Applications

12.1 For Systems Theory

The principle contributes to systems theory by identifying a universal architectural feature that has been recognized in specific domains but not formulated generally. It suggests that boundary maintenance—a core concept in systems thinking—is implemented through a specific mechanism with identifiable properties.

12.2 For Artificial Intelligence and Machine Learning

AI systems increasingly face the challenge of filtering inputs from uncontrolled environments. The principle suggests that robust AI systems will require pre-integrative rejection mechanisms analogous to those in biological and social systems. Current approaches to input validation, adversarial filtering, and out-of-distribution detection can be understood as attempts to implement such mechanisms.

The principle also suggests failure modes to anticipate: AI systems with poorly calibrated rejection may either become overly restrictive (rejecting valid inputs) or overly permissive (accepting adversarial or corrupted inputs).

12.3 For Organizational Design

Organizations must balance openness to new ideas against protection of core functions. The principle provides a framework for analyzing organizational filtering mechanisms and diagnosing dysfunctions. Organizations that reject too readily become rigid and fail to adapt; organizations that accept too readily lose coherence and may integrate elements that undermine their mission.

12.4 For Understanding Social Conflict

Much social conflict involves disputes over rejection thresholds. Groups with different threshold settings will disagree about what should be acceptable for consideration, producing culture wars, generational conflicts, and ideological polarization. The principle suggests that such conflicts are not merely about specific content but about the meta-level question of how calibrated the rejection filter should be.

12.5 For Individual Psychology

Understanding cognitive rejection as a structural feature rather than a bias to be eliminated changes the therapeutic goal from removing resistance to calibrating it appropriately. The aim is not a mind with no filters but a mind with well-calibrated filters.

12.6 Normative Implications and Limitations

PIRP is a descriptive principle—it describes how self-maintaining systems work, not how they should work. However, readers will inevitably draw normative conclusions, and the framework could be invoked to support very different positions:

- **Conservatives** might cite it to defend traditional filtering mechanisms against reforms that raise θ, arguing that societies need strong rejection mechanisms to maintain coherence
- **Progressives** might cite threshold miscalibration pathologies to argue that current thresholds are set too low, causing harmful exclusion of beneficial inputs
- **Authoritarians** might invoke "system integrity maintenance" to justify censorship, exclusion, and suppression of dissent

The principle itself does not adjudicate between these positions. PIRP tells us that *some* threshold exists and that miscalibration in either direction produces pathologies—but it does not tell us what the optimal θ should be for any particular system. That depends on empirical facts about the distribution of harmful versus beneficial inputs, values about which errors are more costly, and context about the system's current state, environment, and adaptive needs.

This normative neutrality is a feature, not a bug. PIRP provides a common language for discussing threshold calibration across different value systems, potentially facilitating dialogue by making explicit what is at stake: not whether filtering should exist, but how it should be calibrated.

13 Limitations and Future Directions

13.1 Limitations

This paper has several limitations:

1. **Abstraction level:** The principle operates at a high level of abstraction. Connecting it to specific mechanisms in each domain requires substantial additional work.

2. **Measurement challenges:** The threshold parameter θ is not directly measurable in most domains. Developing operational measures is a priority for future research.

3. **Interaction complexity:** The nested and interacting nature of rejection filters across levels is acknowledged but not fully analyzed.

4. **Normative questions:** The principle describes what is but does not settle normative questions about what threshold settings should be.

5. **Status as synthesis:** PIRP is best characterized as original theoretical synthesis rather than empirical discovery.

13.2 Future Directions

Several directions for future research emerge:

1. **Domain-specific operationalization:** Developing precise measures of rejection thresholds in specific domains

2. **Cross-level dynamics:** Analyzing how rejection filters at different organizational levels interact

3. **Intervention design:** Developing methods for recalibrating dysfunctional rejection thresholds

4. **Computational modeling:** Building agent-based or dynamical systems models that demonstrate the principle's predictions

5. **Historical analysis:** Examining how rejection thresholds have shifted in historical cases of adaptation and maladaptation

6. **Chemical Organization Theory connections:** Formalizing the relationship between PIRP and the mathematical frameworks developed in Chemical Organization Theory

7. **Comparative empirical studies:** Testing whether the operational criteria identify the same mechanism across domains

8. **AI safety applications:** Developing design principles for AI systems based on PIRP insights

14 Conclusion

The Pre-Integrative Rejection Principle identifies a universal architectural feature of self-maintaining systems: a rapid, low-cost mechanism that excludes potentially incompatible inputs before full analytical processing occurs. This mechanism is not a bias or error but a necessary component of viable systems operating in complex environments.

The principle unifies observations across biology, immunology, ecology, computation, culture, and cognition. These diverse domains exhibit structurally equivalent rejection mechanisms—not in their physical implementation but in their abstract functional architecture.

Understanding this mechanism has practical implications. It suggests that system resilience depends on threshold calibration, that adaptation occurs through parameter adjustment rather than architectural change, and that characteristic pathologies arise from miscalibration. It provides a framework for analyzing conflicts over what should be acceptable for consideration and for designing systems that balance openness with integrity preservation.

Perhaps most importantly, recognizing pre-integrative rejection as a structural feature reframes how we think about resistance, filtering, and exclusion. These are not merely obstacles to understanding or integration but necessary functions that enable systems to maintain themselves in environments containing both nourishment and poison. The question is not whether to reject but how to reject wisely.

A Formal Models of Threshold Dynamics

This appendix presents two formal approaches to modeling the dynamics of the rejection threshold θ: a reinforcement learning model and a Bayesian updating model.

A.1 Reinforcement Learning Model of Threshold Adjustment

Consider a system that adjusts its rejection threshold based on the outcomes of past decisions. Let:

- θ_t = threshold at time t
- x_t = input encountered at time t
- $\phi(x_t)$ = compatibility score of input x_t
- o_t = outcome of decision at time t

The decision rule at each time step is:

$$R(x_t) = \begin{cases} 0 \text{ (reject)}, & \text{if } \phi(x_t) < \theta_t \\ 1 \text{ (accept for analysis)}, & \text{if } \phi(x_t) \geq \theta_t \end{cases} \tag{6}$$

The threshold update rule follows a Rescorla-Wagner-style learning equation:

$$\theta_{t+1} = \theta_t + \alpha \cdot \delta_t \tag{7}$$

where α is a learning rate and δ_t is a signed error signal:

$$\delta_t = \begin{cases} +\epsilon_1, & \text{if false positive (integration failure)} \\ -\epsilon_2, & \text{if false negative (missed opportunity)} \\ 0, & \text{if correct decision} \end{cases} \tag{8}$$

The asymmetry between ϵ_1 and ϵ_2 captures the conservative bias: if $\epsilon_1 > \epsilon_2$, the system raises its threshold more sharply after integration failures than it lowers it after missed opportunities.

Predictions from the RL model:

1. Systems with higher ϵ_1/ϵ_2 ratios will converge to higher (more conservative) thresholds
2. Learning rate α determines adaptation speed
3. Environmental change will cause threshold tracking with lag proportional to $1/\alpha$

A.2 Bayesian Model of Pattern Library Updating

An alternative formalization treats the rejection mechanism as maintaining a probabilistic model of input compatibility. Let:

- H_c = hypothesis that input x is compatible
- H_i = hypothesis that input x is incompatible
- $\phi(x)$ = feature vector of input x
- $P(H_c|\phi(x))$ = posterior probability of compatibility given features

The rejection decision becomes:

$$R(x) = \begin{cases} 1, & \text{if } \frac{P(H_c|\phi(x))}{P(H_i|\phi(x))} > \tau \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where τ is a likelihood ratio threshold that incorporates the cost asymmetry:

$$\tau = \frac{C_{FP}}{C_{FN}} \cdot \frac{P(H_i)}{P(H_c)} \tag{10}$$

Here C_{FP} is the cost of false positives and C_{FN} is the cost of false negatives. When $C_{FP} > C_{FN}$, τ increases, requiring stronger evidence of compatibility before acceptance.

Predictions from the Bayesian model:

1. Systems in stable environments will develop peaked, confident likelihood functions
2. Novel inputs (low probability under both $P(\phi|H_c)$ and $P(\phi|H_i)$) will be rejected by default—the model predicts neophobia
3. The threshold τ can be dynamically adjusted based on context

A.3 Multi-Level Threshold Dynamics

The nested architecture can be formalized by considering multiple interacting thresholds $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(n)}$ at different organizational levels. Cross-level interactions introduce coupling terms:

$$\theta_{t+1}^{(k)} = \theta_t^{(k)} + \alpha_k \cdot \delta_t^{(k)} + \sum_{j \neq k} \gamma_{jk} \cdot (\theta_t^{(j)} - \theta_t^{(k)}) \tag{11}$$

The coupling coefficients γ_{jk} represent how strongly level k's threshold is pulled toward level j's threshold. This captures phenomena such as social pressure shifting individual cognitive thresholds and physiological state affecting cognitive thresholds.

When coupling is strong, thresholds at different levels converge; when coupling is weak, levels can maintain divergent thresholds—capturing cases of professional override where individuals operate with different thresholds than their surrounding culture.

References

Argyris, C. and Schön, D. A. (1978). *Organizational Learning: A Theory of Action Perspective*. Addison-Wesley.

Ashby, W. R. (1956). *An Introduction to Cybernetics*. Chapman & Hall.

Beer, S. (1972). *Brain of the Firm*. Allen Lane.

Burnet, F. M. (1959). *The Clonal Selection Theory of Acquired Immunity*. Vanderbilt University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204.

Conant, R. C. and Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, 1(2):89–97.

Dittrich, P., Ziegler, J., and Banzhaf, W. (2001). Artificial chemistries—a review. *Artificial Life*, 7(3):225–275.

Douglas, M. (1966). *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*. Routledge & Kegan Paul.

Fontana, W. and Buss, L. W. (1994). "The arrival of the fittest": Toward a theory of biological organization. *Bulletin of Mathematical Biology*, 56(1):1–64.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., et al. (2011). Differences between tight and loose cultures: a 33-nation study. *Science*, 332(6033):1100–1104.

Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Wiley.

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74(6):1464–1480.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108(4):814–834.

Hirschman, A. O. (1970). *Exit, Voice, and Loyalty*. Harvard University Press.

Janeway, C. A. and Medzhitov, R. (2002). Innate immune recognition. *Annual Review of Immunology*, 20:197–216.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kreyssig, P., Escuela, G., Reynaert, B., Veloz, T., Ibrahim, B., and Dittrich, P. (2012). Cycles and the qualitative evolution of chemical systems. *PLoS ONE*, 7(10):e45772.

Kriger, B. (2026a). The evolutionary inevitability of predictive processing: a physical constraint argument. *Zenodo*. https://doi.org/10.5281/zenodo.18444910

Kriger, B. (2026b). Non-participation and early exclusion as stability-preserving institutional strategies. *Zenodo*. https://doi.org/10.5281/zenodo.18135188

Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel.

Medzhitov, R. and Janeway, C. A. (1997). Innate immunity: the virtues of a nonclonal system of recognition. *Cell*, 91(3):295–298.

Melnikoff, D. E. and Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, 22(4):280–293.

North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press.

Ostrom, E. (1990). *Governing the Commons*. Cambridge University Press.

Paxton, J. M. and Greene, J. D. (2010). Moral reasoning: hints and allegations. *Topics in Cognitive Science*, 2(3):511–527.

Pettigrew, T. F. and Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5):751–783.

Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review*, 94:251–267.

Rozin, P. and Fallon, A. E. (1987). A perspective on disgust. *Psychological Review*, 94(1):23–41.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6):467–482.

Sydow, J., Schreyögg, G., and Koch, J. (2009). Organizational path dependence. *Organization Studies*, 30:689–721.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

Williamson, O. E. (1985). *The Economic Institutions of Capitalism*. Free Press.

Wolpert, D. M. and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, 3:1212–1217.