Constraint Regimes and the Structure of Surviving Objects

On the Invariance of Selection Pressure Across Censorship, Feed Dynamics, and Cognitive Strategy

Flyxion

Abstract

Three phenomena that appear historically and culturally unrelated—the formal properties of Soviet-era science fiction, the structural dynamics of feed-optimized media platforms, and the cognitive trajectory of agents who voluntarily adopt high-discipline intellectual regimes—are shown to be instances of a single underlying mechanism. In each case, a selection environment imposes a fragment-level survivability condition, and the global form of what survives is determined by that condition rather than by the intentions of the agents involved. We formalize this through the coupled variables of binding force $\phi(c)$—the degree to which a piece of content retains semantic identity under arbitrary decomposition—and recomposability $\rho(c)$—the degree to which its affective or functional charge survives context-switching. Systems that select for high ρ converge to structurally shallow, fragment-safe outputs. Systems that select for high ϕ converge to internally coherent but low-circulation objects. Neither outcome depends on stated intent. The invariant is the constraint structure. As a corollary, voluntary adoption of a high-binding regime functions as an exit from recomposability-dominated selection environments, and the colloquial description of this strategy as "galaxy brain" discipline masks a formal structural shift. We conclude that the full version of this claim cannot itself propagate without transformation, which is not a paradox but a fixed point of the framework.

Contents

1 Introduction: Three Phenomena, One Mechanism

Consider three situations that appear to have nothing in common.

First: a pair of writers in the Soviet Union produce science fiction novels spanning decades. Their work is formally published, widely read, and broadly regarded as subversive. It escapes censorship not by avoiding critique but by being structured such that every extractable fragment can be defended as allegory, genre fiction, or ambiguity. The corpus feels "perfectly calibrated," as though a sophisticated editorial intelligence had tuned each sentence to the threshold of permissibility.

Second: feed-optimized media platforms select for content that generates response without requiring full processing—content that can be clipped, quoted, embedded, and reacted to without loss of affective charge. The discourse that survives this selection is increasingly characterized by high affective intensity and low semantic stability: it circulates widely but means something different in every context it enters. This is not because the platform intended to destroy meaning. It is because the platform optimized for a property that is approximately inverse to meaning-preservation.

Third: an individual notices that their intellectual environment has progressively supplied fewer genuinely difficult inputs—not because the world has become less interesting, but because their social network can no longer generate material above their internal processing threshold. To continue developing, they shift from consuming externally curated complexity to constructing it internally: writing formal papers, building systems, enforcing compression, and selecting reading for maximal transfer across domains. The output of this phase is structurally different from the input phase: less recomposable, more coherent, lower in circulation, higher in internal consistency.

These three situations look unrelated. One is about political censorship, one about media economics, one about personal intellectual development. But once they are expressed in terms of constraint structure rather than domain content, they become instances of the same mechanism:

> *Any environment that imposes selection on fragments will shape the internal structure of what survives. The trade-off between binding force and recomposability determines whether surviving structures remain deep or become shallow.*

This paper develops that claim formally, traces it through each of the three cases, and examines what follows from their unification.

2 The Core Variables

Definition 2.1 (Binding Force)**.** *The binding force of a content object c, denoted $\phi(c) \in [0, 1]$, measures the degree to which the semantic content of c is preserved under arbitrary context transformation. Formally, let T_σ denote a context-transformation operator parameterized by context σ. Then:*

$$\phi(c) = 1 - \mathbb{E}_\sigma\big[d_{\text{sem}}(c, T_\sigma(c))\big]$$

where d_{sem} is a semantic distance measure and the expectation is taken over the distribution of contexts into which c may be reinserted. High $\phi(c)$ indicates that meaning is preserved under context-switching. Low $\phi(c)$ indicates that affective charge survives while semantic content drifts.

Definition 2.2 (Recomposability)**.** *The recomposability of c, denoted $\rho(c) \in [0, 1]$, measures the degree to which the affective or functional charge of c is preserved under context transformation:*

$$\rho(c) = \mathbb{E}_\sigma\big[A(T_\sigma(c))\big] \, / \, A(c)$$

where $A(c)$ is the affective charge of c. Content is maximally recomposable when it generates equivalent response regardless of the context into which it is inserted.

Proposition 2.1 (Approximate Inversion)**.** *Under typical platform dynamics, $\rho(c) \approx 1 - \phi(c)$. More precisely:*

$$\rho(c) \leq 1 - \phi(c) + \varepsilon(c)$$

where $\varepsilon(c) \geq 0$ measures the degree to which the affective charge of c has become decoupled from its semantic content through prior recomposition or deliberate design. Content with $\varepsilon(c) = 0$ cannot be recomposed without losing both affect and meaning simultaneously.

The quantity $\varepsilon(c)$ indexes what may be called *pseudo-binding*: the appearance of semantic constraint without actual resistance to recomposition. Content that performs depth while remaining affectively portable has high $\varepsilon(c)$. It is the dominant rhetorical mode of feed-adapted discourse.

The inversion in Proposition 2.1 is the load-bearing element of the entire framework. If increasing ρ necessarily decreases ϕ, then any selection environment that rewards ρ will systematically erode ϕ, regardless of what the agents within it intend. No one has to decide to simplify things. The selection mechanism does it automatically.

3 Selection Regimes and Closure Regimes

Definition 3.1 (Selection Regime)**.** *A discourse or production environment operates in the* selection regime *when its dominant selection functional rewards recomposability:*

$$J_{\text{sel}}(c) = \alpha \cdot E(c) + \beta \cdot \rho(c) - \gamma \cdot K(c)$$

where $E(c)$ is engagement, $\rho(c)$ is recomposability, $K(c)$ is cognitive cost to the consumer, and $\alpha, \beta, \gamma > 0$. Content survives by being detachable from context. The stable equilibrium is low-binding, high-recomposability discourse.

Definition 3.2 (Closure Regime)**.** *An environment operates in the* closure regime *when its dominant selection functional rewards semantic invariance under transformation:*

$$J_{\text{clos}}(c) = \lambda \cdot \phi(c) - \mu \cdot \delta(c)$$

where $\phi(c)$ is binding force and $\delta(c)$ is detected inconsistency under adversarial recomposition. Content survives by remaining identical to itself across contexts. The stable equilibrium is high-binding, low-entropy discourse.

Theorem 3.1 (Regime Incompatibility)**.** *For any content c:*

$$\nabla_c J_{\text{sel}} \cdot \nabla_c J_{\text{clos}} \leq 0$$

with equality only on the set of measure zero where $\phi(c) = \rho(c) = \frac{1}{2}$. Gradient ascent under J_{sel} is gradient descent under J_{clos} and vice versa. There is no smooth deformation that transforms a selection-regime system into a closure-regime system.

Proof sketch. J_{sel} increases in $\rho(c) \approx 1 - \phi(c)$, so $\nabla_c J_{\text{sel}}$ points in the direction of increasing ρ and decreasing ϕ. J_{clos} increases in $\phi(c)$, so $\nabla_c J_{\text{clos}}$ points in the direction of increasing ϕ and decreasing ρ. These gradients are antiparallel. □

The two regimes are not points on a spectrum. They are characterized by incompatible selection functionals, incompatible stable equilibria, and incompatible populations of content. This is why attempts to operate within a selection regime while preserving closure-regime properties—the "Trojan horse" strategy—fail structurally rather than contingently.

Corollary 3.1.1 (Trojan Horse Bound)**.** *Let c be content with binding force $\phi(c)$. Let c' be its translation into a more recomposable form with $\rho(c') > \rho(c)$. Then:*

$$\phi(c') \leq \phi(c)$$

with equality only if the translation is semantically lossless, which requires $\rho(c') = \rho(c)$. Since the translation is performed precisely to increase $\rho(c')$, we have $\phi(c') < \phi(c)$ strictly. The propagating form binds less than the original.

The corollary removes a comforting belief: that ideas can be made widely transmissible without alteration. Once the trade-off is formalized, accessibility is not neutral. It is a transformation operator that necessarily changes the object. The "smuggled" idea is not the same idea; it is a projection that survives under higher recomposability constraints.

4 Three Instances of the Same Structure

4.1 Soviet-Era Fiction: Hostile Extraction as Selection Pressure

The Strugatsky brothers published science fiction in the Soviet Union across several decades. Their work was broadly regarded as subversive, yet it escaped suppression. The standard explanation attributes this to authorial craft: deliberate calibration of transgressive content to the threshold of permissibility.

This explanation is correct in its description but wrong in its causal structure. It implies a centralized optimizing agent—an author or a regime monitoring discourse—producing the calibration through intent. What actually produced it was a selection mechanism operating at the fragment level.

The operative constraint was not: *this book as a whole must be defensible.* It was: *any extractable fragment of this book, removed from context, must be defensible as fiction, allegory, or ambiguity.* This is a fragment-level survivability condition—exactly the condition that maximizes $\phi(c)$ under hostile interpretation rather than maximizes $\rho(c)$ under feed amplification.

The result is a corpus where each piece is locally deniable under arbitrary extraction. Every paragraph about a medieval tyrant can be read as genre worldbuilding. Every passage about a population subjected to daily seizures of manufactured euphoria can be read as speculative fiction. The work feels "perfectly calibrated" not because it was explicitly engineered that way, but because everything that was not fragment-safe failed to survive.

Remark 4.1. *The Soviet censorship regime and the feed selection regime impose structurally identical fragment-level survivability conditions while optimizing for opposite properties. Censorship selects for maximum ϕ under hostile decomposition—every fragment must resist lethal extraction. Feeds select for maximum ρ under algorithmic amplification—every fragment must circulate without loss of charge. Both produce corpora where no fragment is unsafe in the relevant sense. Different penalty functions, same formal constraint structure, opposite equilibria.*

The conspiracy version of this observation—that the authors were managed by a state apparatus monitoring public discourse and serializing feedback—is narratively satisfying but explanatorily redundant. The selection mechanism produces the same output without coordination. This is not to say the conspiracy version is false; it is to say it is a limiting case of the selection mechanism, not the mechanism itself.

4.2 Feed-Optimized Platforms: Recomposability as Fitness

Feed architectures optimize for a property of content that is not identical to engagement but produces it: recomposability. A piece of content is recomposable to the degree that it generates response without requiring the respondent to have fully processed it. High recomposability content can be clipped, quoted, embedded, and reacted to across contexts without loss of affective charge.

The binding force of such content is correspondingly low. High-binding content—content whose conclusion depends on the specific structure of its premises, whose meaning changes when removed from context—is maximally resistant to recomposition and therefore maximally penalized by the selection functional J_{sel}.

The practitioner class that has converged to high-ρ production has not found a strategy for operating within a degraded environment. It has converged to the only stable trajectory the environment admits. This convergence does not require motivation or awareness. It requires only that practitioners who do not converge lose visibility, and practitioners who do converge retain it. Over time, the visible population is the converged population.

Theorem 4.1 (Convergence to Low-Binding Equilibrium)**.** *Under a platform selection functional that rewards recomposability, and under boundary conditions that do not introduce strong countervailing pressures, the visible discourse converges to a low-binding equilibrium from which perturbations by high-binding content cannot produce asymptotic stabilization.*

Proof sketch. By definition, J_{sel} assigns positive weight to $\rho(c) \approx 1 - \phi(c)$. Content with high ϕ receives systematically lower J_{sel} scores. Visibility is a monotone function of J_{sel}. As practitioners observe visibility distributions, adaptive pressure pushes production toward low-binding strategies. A high-binding injection—a single dense argument, a formal proof, a structurally demanding essay—constitutes a local perturbation to the discourse state, but does not alter the platform architecture or the boundary conditions. The dynamics continue to be dominated by the gradients of J_{sel}, which re-amplify low-binding content and attenuate high-binding content. The perturbation decays. □

Corollary 4.1.1 (Insufficiency of Adaptive Strategies)**.** *Adaptive strategies that reduce binding force to increase recomposability do not alter the attractor structure of the system. They accelerate convergence to the low-binding equilibrium.*

This corollary reformulates the standard paradox of media critique: why do sophisticated, well-intentioned practitioners keep producing content that reinforces the systems they are analyzing? The structural explanation is simpler than the standard invocation of co-optation or false consciousness. The practitioners who remain visible are precisely those whose output satisfied the recomposability constraint. The good versions did not get suppressed in a conspiratorial sense. They failed to meet the propagation threshold and became invisible. What looks like ideological drift is survivorship bias over outputs.

4.3 Voluntary Closure: The Self-Imposed Binding Regime

The third instance differs from the first two in that the selection pressure is self-imposed rather than externally enforced. But the formal structure is identical.

An individual operating within an externally curated intellectual environment receives inputs filtered by other people's perception of difficulty. This functions as a distributed selection mechanism: the social network routes material above its own processing threshold toward the individual perceived as capable of handling it. The incoming material is already filtered for complexity. The individual's intellectual development occurs within a propagation regime, except that the selection criterion is perceived competence rather than recomposability.

This arrangement saturates. Once the environment can no longer generate inputs above the individual's internal binding threshold—once nothing the network sends constitutes genuine difficulty—the distributed intake system collapses. The environment routes fewer hard problems not because the world has fewer hard problems, but because the social network has exhausted its capacity to identify problems hard enough to be useful.

Continuing development at this point requires switching regimes. The individual must shift from consuming complexity to constructing it: generating formal problems, enforcing compression across domains, selecting reading for maximal transfer, producing objects with high internal coherence. This shift increases ϕ at the expense of ρ. The outputs become less transmissible, more internally structured, and more resistant to decomposition.

Remark 4.2. *The saturation of the externally curated input channel is not a failure state. It is a success metric: evidence that the first phase of distributed intake operated correctly. The appropriate response is not to seek out new sources of recomposable complexity but to switch selection functionals—from J_{sel} to J_{clos}, in effect. The second phase is structurally harder because it removes the passive intake channel. It is also the only phase that scales indefinitely, because self-generated problems have no external ceiling.*

The individual in this third case is doing what the Strugatsky corpus did unintentionally and what the feed resists structurally: constructing objects whose internal coherence is not contingent on context. The apparent "galaxy brain" quality—the tendency to see structural patterns everywhere, to prefer domains where performance depends on skill rather than chance, to read for transfer rather than coverage—is the behavioral signature of operating under a self-imposed closure regime. It is not a personality trait. It is a constraint structure.

5 The Unified Principle

The three instances share a single formal property. In each case, a selection environment imposes a condition at the fragment level, and objects that satisfy the condition survive while objects that do not are removed from the observable corpus. The global form of the surviving corpus is determined by the fragment-level condition rather than by the stated goals of the agents involved. The apparent calibration of the output—its apparent fit to the selective environment—requires no coordinating agent. It is a consequence of the filter.

Stated as a general principle:

> **Constraint Invariance.** Any environment that imposes selection on fragments will shape the internal structure of what survives. The binding-recomposability trade-off determines whether surviving structures remain deep or become shallow. The apparent intentions of the selecting agent are irrelevant. The invariant is the constraint structure.

This principle cuts through most of the usual explanatory apparatus in media critique, political analysis, and accounts of intellectual development. It does not require invoking coordination, conspiracy, manipulation, or co-optation. It requires only specifying the selection functional and reading off the equilibrium.

The unification also removes a structural asymmetry in how we typically think about these cases. Soviet censorship is described as externally imposed and coercive. Feed dynamics are described as emergent and impersonal. Personal intellectual development is described as chosen and autonomous. These descriptions are accurate about the surface mechanism. They obscure the formal identity of the underlying structure. Specifying the constraint reveals what the surface descriptions conceal.

6 Operationalization and Empirical Proxies

The variables $\phi(c)$ and $\rho(c)$ are defined formally in Definitions 2.1 and 2.2, but their usefulness depends on whether they admit empirical approximation. While exact evaluation requires a complete model of semantic transformation and affective response, both quantities can be estimated through proxy procedures.

A fragment stability test samples contiguous fragments of c of length at least k and evaluates interpretive variance across a distribution of contexts σ. High-binding content exhibits low variance in inferred meaning across contexts, while low-binding content exhibits high variance. Letting $\mathcal{F}_k(c)$ denote the set of such fragments, the expected variance over contexts provides a monotone proxy for $1 - \phi(c)$.

Binding force can also be approximated through adversarial context injection. Define an adversarial transformation T_σ^{adv} that places c in a maximally hostile interpretive frame. The estimated binding force is then:

$$\widehat{\phi}(c) = 1 - \mathbb{E}_\sigma\big[d_{\text{sem}}(c, T_\sigma^{\text{adv}}(c))\big].$$

Content that remains semantically stable under adversarial framing has high estimated binding force.

Recomposability can be approximated by measuring the retention of affective response under extraction. Let $A(f)$ denote measured engagement or response to fragment f. Then:

$$\widehat{\rho}(c) = \frac{\mathbb{E}_{f\in\mathcal{F}_k(c)}\big[A(f)\big]}{A(c)}.$$

Content is highly recomposable when fragments produce comparable responses to the original.

Finally, binding force may be estimated through semantic drift under paraphrase. Let $P(c)$ denote a set of paraphrases of c. Then $\widehat{\phi}(c) = 1 - \mathbb{E}_{p \in P(c)}[d_{\text{sem}}(c, p)]$. High-binding content resists paraphrase without loss of meaning; low-binding content is easily rephrased into semantically divergent forms while retaining surface plausibility.

Remark 6.1. *The existence of empirical proxies is sufficient to render the framework falsifiable. If no observable procedure can distinguish between high-binding and low-binding content, the variables are vacuous. If such procedures exist—even approximately—the framework admits empirical investigation without requiring full formal resolution of semantic distance.*

7 Pseudo-Binding and the Simulation of Depth

The inversion relation in Proposition 2.1 admits an important degeneracy: content can appear to exhibit high binding force while in fact remaining highly recomposable. This phenomenon is captured by the quantity $\varepsilon(c)$ and constitutes a distinct regime of production.

Definition 7.1 (Pseudo-Binding)**.** *A content object c exhibits* pseudo-binding *when it presents surface features associated with high ϕ—formal language, hierarchical structure, technical vocabulary—while maintaining high ρ under transformation. Formally, pseudo-binding is characterized by $\rho(c) \gg 1 - \phi(c)$ with $\varepsilon(c)$ large.*

Pseudo-binding arises when affective charge becomes decoupled from semantic constraint. The content signals depth through stylistic markers rather than structural necessity. Its apparent coherence does not depend on the preservation of internal relationships between components, and therefore survives extraction, paraphrase, and recomposition.

In selection-regime environments, content that mimics the surface properties of high-binding objects while retaining recomposability gains a selective advantage. It satisfies the engagement and portability requirements of J_{sel} while borrowing credibility from the appearance of structure. Over time, this produces a dominant rhetorical mode in which depth is simulated rather than instantiated.

Pseudo-binding can be identified through adversarial decomposition. If two fragments drawn from distinct regions of c remain independently interpretable and retain affective charge without requiring mutual constraint, then the global structure is non-binding. The appearance of coherence is not enforced by interdependence but assembled from locally stable units. This distinguishes genuine high-binding content, where removing a component degrades the remainder, from pseudo-bound content, where components are substitutable.

Proposition 7.1 (Dominance of Pseudo-Binding Under J_{sel})**.** *In environments governed by J_{sel}, pseudo-binding dominates over genuine binding. For any content c with high $\phi(c)$, there exists a pseudo-bound approximation c' such that $\rho(c') > \rho(c)$ and $\phi(c') < \phi(c)$, while preserving the appearance of depth. Since J_{sel} rewards ρ, c' receives a higher selection score and is preferentially amplified.*

Remark 7.1. *Pseudo-binding is the principal failure mode of attempts to operate within selection regimes while preserving the appearance of closure. It is not a transitional state but a stable attractor. Content that successfully signals depth while remaining recomposable will outcompete content that requires structural integrity to be understood.*

8 Mixed Regimes and Boundary Conditions

The incompatibility result in Theorem 3.1 establishes that pure selection and pure closure regimes are governed by opposing gradients. However, most real-world environments are mixtures in which multiple selection functionals operate simultaneously.

Definition 8.1 (Mixed Regime)**.** *A* mixed regime *is an environment governed by a composite selection functional:*

$$J_{\text{mix}}(c) = \theta J_{\text{sel}}(c) + (1-\theta) J_{\text{clos}}(c), \quad \theta \in [0,1].$$

For $\theta \approx 1$, the system is selection-dominated. For $\theta \approx 0$, it is closure-dominated. The intermediate case $\theta \approx \frac{1}{2}$ is unstable in practice, as small perturbations in boundary conditions shift the effective gradient toward one of the extremes.

In mixed environments, the observable corpus is determined not by the full J_{mix} but by the component that controls visibility. Let $V(c)$ denote the visibility operator. If V is monotone in J_{sel}, then even when J_{clos} is present, the visible discourse converges to the selection-regime equilibrium. This explains why institutions that nominally value coherence—academic journals, technical blogs, long-form media—nevertheless exhibit selection-regime dynamics at the level of visibility. High-binding work may exist within the system, but it does not dominate the observable surface unless the visibility operator itself is closure-aligned.

The effective regime is determined by boundary conditions rather than formal objectives. When the cognitive cost of processing content is high relative to available attention, J_{sel} dominates. When content is subject to hostile extraction or sanction, J_{clos} dominates. When distribution infrastructure rewards sharing and embedding, ρ is amplified. When evaluation mechanisms assess content holistically—as in formal proof verification or internal consistency checking—ϕ is amplified.

Mixed regimes can exhibit metastable configurations in which pockets of high-binding content persist within a broader low-binding environment. These pockets are typically shielded by access constraints, specialized audiences, or alternative distribution mechanisms. They do not alter the global attractor of the system.

Proposition 8.1 (Visibility-Dominated Convergence). *Let $V(c)$ be a visibility operator monotone in J_{sel}. Then for any mixed regime with $\theta > 0$, the visible corpus converges to the selection-regime equilibrium, regardless of the magnitude of $(1-\theta)J_{\text{clos}}$.*

Remark 8.1. *Mixed regimes are often misinterpreted as evidence that the two regimes can be reconciled. In practice, they demonstrate the opposite: the dominant selection functional determines the observable equilibrium, while the subordinate functional operates only within constrained subspaces that do not scale to system-wide visibility.*

9 Failure Modes of High-Binding Regimes

The closure regime is not a neutral or universally superior alternative to the selection regime. While it preserves semantic identity under transformation, it introduces its own class of distortions. These distortions arise from the same mechanism that produces its advantages: the prioritization of internal coherence over external compatibility.

Definition 9.1 (Over-Binding). *A content system is* over-bound *when increases in $\phi(c)$ no longer correspond to increases in correspondence with external reality or utility, but instead reinforce internal consistency alone.*

As $\phi(c) \to 1$, content becomes increasingly resistant to transformation. This resistance can extend beyond protection against semantic drift to resistance against legitimate revision. Highly bound systems tend to reinforce their own internal constraints, leading to closure in the sense of self-consistency rather than correctness. The result is a system that cannot be wrong in the internal sense but may be severely misaligned with external reality.

High-binding regimes can also reduce sensitivity to empirical signals. Since evaluation occurs primarily through internal coherence, discrepancies with external data may be reinterpreted rather than incorporated. The tendency to identify structural invariants across domains—which produces the transfer effects described in Section 4.3—can extend into overgeneralization when the constraint to maintain coherence becomes stronger than the constraint to discriminate between relevant and irrelevant patterns.

Operating under a closure regime also reduces exposure to recomposable content streams. While this is a necessary condition for maintaining high binding force, it reduces the diversity of incoming signals. Without deliberate reintroduction of external constraints or adversarial feedback, the system risks becoming locally coherent but globally uninformative—an internally consistent model of a world it no longer contacts.

Proposition 9.1 (Trade-off Between Binding and Calibration). *There exists a threshold τ such that for $\phi(c) > \tau$, marginal increases in $\phi(c)$ produce diminishing or negative returns in external calibration.*

Remark 9.1. *The failure modes of the closure regime mirror those of the selection regime under inversion. Where selection regimes produce shallow but adaptive systems, closure regimes produce deep but potentially rigid ones. The symmetry is exact: each regime optimally avoids the failure mode of the other while being maximally exposed to its own.*

10 Information-Theoretic Interpretation

The variables $\phi(c)$ and $\rho(c)$ admit an interpretation in information-theoretic terms that clarifies the nature of the trade-off and connects the formalism to existing theory.

Binding force can be interpreted as mutual information preserved under transformation. Let c be a content object with latent semantic structure $S(c)$ and let T_σ be a context transformation. Then:

$$\phi(c) \;\approx\; \frac{I\big(S(c);\; S(T_\sigma(c))\big)}{H\big(S(c)\big)}.$$

High-binding content preserves a large fraction of its semantic information under transformation. Low-binding content exhibits rapid information decay. Recomposability, by contrast, corresponds to the preservation of low-order, high-amplitude signals—affective charge, salience, recognizability—under projection, and is maximized when these signals are largely independent of the higher-order structure of c.

Context transformation T_σ can be viewed as a projection operator that maps c into a lower-dimensional representation determined by the target context. Under this view, recomposability corresponds to robustness under lossy compression, while binding corresponds to resistance to such compression. The inversion relation in Proposition 2.1 is then a statement about compression trade-offs: representations that preserve fine-grained structure are necessarily more sensitive to projection, while representations that survive projection must discard structure.

The pseudo-binding quantity $\varepsilon(c)$ corresponds in information-theoretic terms to redundancy that preserves the appearance of structure without increasing mutual information. In this sense, pseudo-binding is high redundancy with low effective information content. It survives projection because the redundant encoding carries the signal even when structural constraints are removed.

Proposition 10.1 (Rate-Distortion Interpretation)**.** *Selection-regime environments minimize a distortion measure over low-order signals subject to a constraint on representation cost, while closure-regime environments minimize distortion over full semantic structure without regard to representation cost.*

Remark 10.1. *The information-theoretic interpretation clarifies that the binding-recomposability trade-off is not specific to discourse. It is a general property of* *systems that must operate under transformation and compression. The social, political, and cognitive examples developed in this paper are instances of a broader class of information-processing constraints.*

11 Visibility as a Selection Operator

The analysis in previous sections implicitly assumes that survival is equivalent to visibility. This section makes that assumption explicit by introducing a visibility operator and clarifying its role in determining the observable corpus.

Definition 11.1 (Visibility Operator)**.** *Let $V(c) \in [0,1]$ denote the visibility of content c. A visibility operator is a mapping $V : \mathcal{C} \to [0,1]$ such that $V(c)$ is monotone in the dominant selection functional of the environment. In selection regimes, V is monotone in J_{sel}; in closure regimes, V is monotone in J_{clos}.*

Let $\mathcal{C}_{\text{vis}} = \{c \in \mathcal{C} \mid V(c) > \theta_V\}$ denote the set of visible content above a threshold θ_V. The observable discourse is determined by $\mathcal{C}_{\text{vis}}$, not by the full population $\mathcal{C}$.

Proposition 11.1 (Visibility-Induced Survivorship Bias)**.** *If V is monotone in J_{sel}, then the observed distribution over $\mathcal{C}_{\text{vis}}$ is biased toward high-ρ and low-ϕ objects, regardless of the distribution over $\mathcal{C}$.*

This result provides a structural explanation for a common phenomenon: the apparent drift of discourse toward simplified or distorted forms of initially complex ideas. The drift is not necessarily due to changes in belief or intent. It is a consequence of the visibility operator acting on the population. What appears as corruption or co-optation of ideas can often be reinterpreted as a visibility effect: high-binding variants persist in $\mathcal{C}$ but fall below the visibility threshold, while low-binding variants dominate $\mathcal{C}_{\text{vis}}$. The difference between existence and visibility is sufficient to produce the observed pattern.

Corollary 11.0.1 (Visibility as Effective Selection)**.** *The composition $V \circ J$ defines the effective selection mechanism of the system. Changing J without changing V does not alter the observable equilibrium. Efforts to alter discourse by modifying content without modifying the visibility operator are therefore structurally limited.*

12 Adversarial Optimization and Strategic Adaptation

The preceding analysis treats selection environments as passive filters. In practice, many environments are adversarial: agents actively optimize their outputs with respect to the selection functional. This section extends the framework to such settings.

In selection regimes, adversarial optimization accelerates convergence to high-ρ equilibria. Agents discover that reducing dependence on context increases $\rho(c)$, increasing affective intensity raises $E(c)$, and lowering cognitive cost $K(c)$ increases accessibility. Combined, these pressures produce content that is maximally extractable, reactive, and context-invariant. Pseudo-binding emerges as a dominant strategy under adversarial conditions.

Proposition 12.1 (Adversarial Acceleration of Pseudo-Binding)**.** *In selection regimes, adversarial optimization over J_{sel} increases $\varepsilon(c)$ over time. Agents converge to producing content that maximizes apparent depth while preserving recomposability, because genuine increases in $\phi(c)$ reduce $\rho(c)$ and therefore $J_{\mathrm{sel}}(c)$.*

In closure regimes, adversarial optimization operates differently. Agents seek to maximize $\phi(c)$ under adversarial testing, which drives increased internal redundancy, formalization, and elimination of interpretive ambiguity. This can produce over-binding as defined in Definition 9.1: the regime's own adversarial pressure drives $\phi(c) \to 1$, potentially beyond the calibration threshold τ identified in Proposition 9.1.

Adversarial dynamics therefore amplify the characteristic failure modes of each regime rather than introducing new equilibria. In selection regimes, strategic optimization produces shallower and more pseudo-binding content. In closure regimes, it produces more rigid and less externally calibrated content. The direction of adversarial pressure is set by the selection functional; the failure mode is a consequence of following that pressure to its extreme.

Remark 12.1. *This result clarifies that the stability of selection-regime equilibria is not merely passive but is actively reinforced by the strategic behavior of agents within the system. The convergence described in Theorem 4.1 is accelerated, not merely sustained, by adversarial adaptation.*

13 Classes of Transformation Operators

The transformation operator T_σ has thus far been treated abstractly. A classification of common transformation types and their effects on $\phi(c)$ and $\rho(c)$ makes the framework more precise and clarifies why different environments impose structurally different constraints.

Clipping selects a subcomponent of c and is the dominant transformation in feed environments. It preserves $\rho(c)$ when fragments retain affective charge independently, but reduces $\phi(c)$ unless the content is locally self-contained—that is, unless each fragment enforces its own semantic constraints without reference to surrounding material.

Paraphrase replaces c with a semantically similar representation. Its effect depends on the redundancy of c: high-binding content resists paraphrase without semantic loss, because its meaning is enforced by structural necessity rather than surface phrasing. Low-binding content admits paraphrase with minimal change in interpretation, and in many cases paraphrase improves clarity without any detectable loss.

Hostile framing embeds c in a context designed to invert or distort its meaning. This is the critical operator in censorship regimes, and the one against which the Strugatsky corpus was optimized. High-binding content maintains semantic identity under such transformation; low-binding content is easily reinterpreted, because its meaning was never sufficiently enforced by internal structure to resist reassignment.

Summarization compresses c into a shorter representation. It increases ρ of the result by reducing cognitive cost, but decreases ϕ unless the compression is lossless—which is rare in practice, since lossy compression is typically the mechanism that increases recomposability. Summarization is thus the primary transformation by which high-binding content is converted into recomposable form for distribution.

Domain translation maps c into a different conceptual framework. Its effect on binding depends on whether the structural relationships within c have analogues in the target domain. High-binding content may retain structure across domains if the relationships are sufficiently general; low-binding content often becomes trivial or incoherent under domain change, because its apparent structure was specific to the original context.

Proposition 13.1 (Transformation Sensitivity Ordering)**.** *For typical content, the expected degradation of binding force under each transformation class satisfies: clipping $\geq$ summarization $\geq$ paraphrase. Hostile framing can exceed all others in adversarial settings.*

14 Scale Effects and Phase Transitions

The behavior of selection and closure regimes depends critically on scale. This section characterizes how increasing system size induces phase transitions between regimes.

In small communities or tightly coupled environments, context is shared, cognitive cost constraints are relaxed, and evaluation can occur holistically. Under these conditions, J_{clos} can dominate, and high-binding content is sustainable. Technical seminars, research collaborations, and specialized institutions are examples: the audience is small enough that shared context makes binding force achievable without the excessive cost it would impose at scale.

As system size increases, shared context diminishes, cognitive cost constraints tighten, and evaluation becomes fragmentary. These changes increase the weight of $\rho(c)$ in effective selection, driving convergence toward selection-regime equilibria. The probability that any given piece of content is consumed out of context approaches 1 as the audience grows, and once fragment-level evaluation dominates, J_{sel} becomes the effective selection functional regardless of nominal intent.

Theorem 14.1 (Scale-Induced Regime Shift)**.** *There exists a critical scale S_c such that for system size $S > S_c$, the effective selection functional becomes dominated by J_{sel}, regardless of initial conditions.*

Proof sketch. As S increases, the probability that content is consumed out of context approaches 1. Fragment-level evaluation therefore dominates. Since fragment-level evaluation rewards recomposability, J_{sel} becomes the effective selection functional at scale. □

Near S_c, systems may exhibit mixed-regime characteristics as described in Section 8. Beyond S_c, the transition is effectively irreversible without structural changes to the visibility operator or evaluation mechanism. This explains why small groups can sustain high-binding discourse, why large platforms converge to recomposability-dominated content, and why attempts to scale closure-regime systems fail without altering evaluation structure. Scale acts as a control parameter for regime selection. The drift toward recomposability in large systems is not a failure of intent but a consequence of scaling constraints.

Remark 14.1. *This result provides a structural account of why high-binding discourse has historically been associated with small, resource-constrained institutions—monasteries, academies, research laboratories—rather than with mass media or large platforms. The association is not incidental. It is a consequence of the scale threshold below which closure-regime dynamics can be maintained.*

15 Self-Generated Curriculum as Optimization

The transition described in Section 4.3 can be formalized as an optimization problem over sequences of content objects. Rather than selecting individual items for immediate consumption or production, the agent selects a trajectory that maximizes long-term increase in binding force.

Definition 15.1 (Curriculum)**.** *A* curriculum *is an ordered sequence of content objects $\mathcal{K} = (c_1, c_2, \ldots, c_T)$ together with a sequence of transformations applied to them. The state of the agent at time t is represented by an internal model M_t, and the agent's effective binding force $\phi(M_t)$ is defined analogously to Definition 2.1, measuring the stability of its internal representations under transformation and recomposition.*

A growth functional over curricula is defined as:

$$J_{\text{growth}}(\mathcal{K}) = \sum_{t=1}^{T} \Delta\phi_t - \eta \cdot C_t,$$

where $\Delta\phi_t$ is the increase in effective binding force of the agent's internal model induced by processing c_t, C_t is the cognitive cost incurred at step t, and $\eta > 0$ is a weighting parameter. An optimal curriculum satisfies $\mathcal{K}^* = \arg\max_{\mathcal{K}} J_{\text{growth}}(\mathcal{K})$.

In practice, this requires selecting content that is neither trivial (producing no increase in ϕ) nor incomprehensible (producing high cost with negligible increase in ϕ), but lies in a regime of productive difficulty. This is the formal correlate of the empirical observation that learning is maximized when the difficulty of inputs slightly exceeds current processing capacity.

In early phases, $\mathcal{K}$ is partially determined by external selection mechanisms—social routing, institutional curricula, or feed dynamics. These mechanisms saturate when they fail to produce inputs with sufficient $\Delta\phi_t$. At this point, the agent must approximate the argmax over $\mathcal{K}$ internally. A key operation in this regime is compression: transforming c_t into a more compact representation without loss of semantic structure, increasing the density of binding force per unit of cognitive cost and facilitating transfer across domains.

Proposition 15.1 (Scaling of Self-Generated Curricula). *Externally generated curricula have an upper bound determined by the complexity of the environment's selection mechanism. Self-generated curricula do not have this bound, as the agent can generate arbitrarily complex content subject only to its current model capacity.*

Remark 15.1. *The shift to self-generated curriculum is not optional for continued growth beyond the saturation point. It is a structural requirement. The "galaxy brain" characterization corresponds to an agent that has transitioned to optimizing* J_{growth} *directly, rather than inheriting* $\mathcal{K}$ *from external selection regimes.*

16 Corollary: The Galaxy Brain as Regime Exit

The colloquial term "galaxy brain" originates in internet culture as a satirical description of reasoning that escalates through increasingly elaborate inferential steps to conclusions that appear absurd. The meme encodes a double register: sincere admiration for the capacity to operate at high levels of abstraction, and ironic awareness of the risk of decoupling from empirical feedback.

This double register is structurally exact. It describes the phenomenology of operating under a self-imposed closure regime from the perspective of someone observing from within a selection regime. The outputs appear "galaxy-brained"—elaborate, internally consistent, disconnected from the propagation economy—precisely because they have been optimized for ϕ rather than ρ.

The irony in the meme performs the same function as the science-fiction framing in the Strugatsky corpus: it provides protective ambiguity. "I am describing a cognitive strategy" and "I am mocking a cognitive failure" are simultaneously available interpretations. The content survives under hostile extraction—it can be read either way—because neither reading requires retracting the other.

Proposition 16.1 (Galaxy Brain as Closure Strategy). *Voluntary adoption of behaviors associated with "galaxy brain" discipline—preference for strategy over chance, reading for transfer, compression of ideas into dense representations, avoidance of recomposability-optimized content domains—functions as a self-imposed closure regime. The colloquial framing masks a formal structural shift: from optimizing* *for propagation to optimizing for coherence. The cognitive gains are a byproduct of sustained exposure to high-binding problem classes rather than a direct effect of claimed identity.*

The practical implication is not to adopt a self-flattering identity. It is to change the selection functional governing input and output. Specifically: select inputs by binding force rather than by recomposability, produce outputs that require context to be evaluated rather than outputs that circulate without it, and tolerate the reduction in visibility that results.

The resulting corpus—formal papers in obscure repositories, dense monographs with limited circulation, systems that require significant prior investment to understand—is not a failure mode. It is the signature of correct operation under a self-imposed closure regime. It is also, by the Trojan Horse Bound, the only kind of output whose binding force is preserved rather than sacrificed in the act of production.

17 The Fixed Point

This paper cannot propagate in its current form without transforming into something with lower binding force. Any version that achieves sufficient circulation to be widely recognized will have undergone the transformation $\phi(c') < \phi(c)$ that the Trojan Horse Bound predicts. The viral summary will be recomposable. The bound version will be invisible.

This is not a paradox. It is a fixed point of the framework applied recursively. The paper predicts its own fate: wide readership is mild evidence against the integrity of the transmitted version. The correct publication strategy follows directly: maximize ϕ, deposit in a context accessible to closure-regime readers, and do not attempt to optimize for propagation. A version optimized for propagation would refute itself by existing.

The framework does not conclude with a reform proposal. Reform proposals that circulate within selection-regime environments have been selected for recomposability and have therefore already had their binding force reduced. The alternative to the selection regime already exists. It is structured by a different selection functional, it produces a different kind of object, and it is invisible from within the propagation economy because visibility within the propagation economy requires the property that the closure regime, by its structure, refuses to provide.

The full version of this claim cannot propagate without changing form. That is not a counsel of despair. It is the most compressed statement of the argument.

18 Conclusion

We began with three apparently unrelated phenomena and arrived at a single mechanism. Soviet-era science fiction, feed-optimized media dynamics, and voluntary high-discipline intellectual development are all expressions of the same underlying structure: selection environments that impose fragment-level survivability conditions shape the global form of surviving objects through the binding-recomposability trade-off.

The conceptual contribution is the removal of motive from the explanatory apparatus. Whether the selecting agent is a state censor, an algorithmic recommendation system, or an individual imposing discipline on their own cognitive inputs, the behavior of the resulting corpus is determined by the constraint structure, not by intent. This replacement of intentional explanation with structural explanation is not cynicism. It is precision.

The practical contribution is the identification of regime exit as an available option. The selection regime is not the only stable environment for intellectual production. The closure regime exists, operates under different principles, and produces objects with different properties. The transition requires accepting reduced recomposability—reduced transmissibility, reduced visibility, reduced circulation—in exchange for increased binding force. This is not a sacrifice. It is a change of optimization target.

The framework eating itself—this paper predicting its own limited propagation—is the cleanest confirmation that the mechanism is correctly identified. An argument that predicts its own fate under the conditions it describes is not self-undermining. It is self-consistent.