

# Scalar Reward Collapse: A General Theory of Eigenstructure Evaporation in Closed-Loop Systems

James Beck  
*Independent Researcher*  
December 2024

---

## Abstract

We establish a general dynamical result for systems governed by scalar reward optimization in closed loops. Given a finite state space  $X$ , a scalar reward function  $r: X \rightarrow \mathbb{R}$ , and a multiplicative reweighting operator  $T$  that updates probability distributions according to reward, we prove three fundamental theorems: (1) **Eigenstructure Evaporation** (Lemma 2.1): all non-reward-maximal modes decay exponentially under repeated application of  $T$ ; (2) **Fixed-Point Convergence** (Theorem 2.5): closed-loop dynamics  $p_{\{t+1\}} = T(p_t)$  converge to a fixed point concentrated on reward-maximal states with reduced entropy; (3) **Irreversibility** (Corollary 2.7): once collapse occurs, the same operator  $T$  cannot restore suppressed modes without exogenous forcing or structural modification. We provide a  $\Delta t$  interpretation showing that scalar optimization forces multi-timescale systems into single-mode temporal collapse. These results supply the mathematical foundation for observed phenomena including platform content homogenization, AI model mode collapse, market herding dynamics, and institutional monoculture formation.

---

## 1. Introduction

### 1.1 Motivation

Across diverse domains—recommendation systems, reinforcement learning, market dynamics, evolutionary selection, institutional optimization—we observe a recurring pathology: systems governed by scalar optimization in closed loops exhibit progressive loss of diversity, convergence to narrow behavioral modes, and resistance to recovery interventions.

Examples include:

- **Platform ecosystems:** Content feeds converge to algorithmically-preferred formats despite diversity initiatives
- **AI training:** RLHF-aligned models exhibit mode collapse and style homogenization
- **Financial markets:** Herding behavior and flash crashes driven by algorithmic trading
- **Institutional dynamics:** Organizations converging to metric-optimized but functionally degraded states

These phenomena share a common structure: a scalar reward signal drives iterative reweighting of a probability distribution over states, creating a feedback loop where high-reward states become increasingly dominant.

This paper establishes the **mathematical necessity** of such collapse, proving that it is not a contingent failure of implementation but an inevitable consequence of the optimization architecture itself.

## 1.2 Main Results

We prove three core theorems about scalar reward optimization in closed loops:

### Theorem 1 (Eigenstructure Evaporation, Lemma 2.1):

Under multiplicative reweighting by scalar reward, all non-maximal modes decay exponentially. The decay rate is determined by the reward gap  $\Delta r = r_{\max} - r_i$ .

### Theorem 2 (Fixed-Point Convergence, Theorem 2.5):

Closed-loop dynamics  $p_{\{t+1\}} = T(p_t)$  converge to a fixed point concentrated on the reward-maximal set  $S_{\max}$ . Any such fixed point has lower entropy than initial distributions with mass outside  $S_{\max}$ .

### Theorem 3 (Irreversibility, Corollary 2.7):

Once the system reaches a collapsed fixed point, continued application of the same operator  $T$  cannot restore previously suppressed modes. Recovery requires either exogenous forcing (external injection of suppressed modes) or structural modification (changing the reward function or operator).

These results are **domain-agnostic**: they apply to any system with the specified structure, regardless of whether the “states” represent content types, model outputs, market strategies, or organizational behaviors.

## 1.3 Relation to $\Delta t$ Framework

This work connects to the  $\Delta t$  framework developed in prior papers [Beck 2024a,b,c], which analyzes multi-timescale systems and their stability properties. Section 6 provides the temporal interpretation: scalar optimization collapses multi-timescale structure by forcing all system dynamics to evolve at a single cadence determined by the optimization update frequency. This produces what we term **temporal monoculture**—the elimination of slow modes that would otherwise provide system adaptability and resilience.

## 1.4 Scope and Structure

This paper presents pure theory. We deliberately exclude domain-specific examples to maintain generality; applications to specific domains are left to future work. Our goal is to establish the mathematical backbone for understanding collapse phenomena across systems.

**Structure:** - Section 2: Formal definitions and setup - Section 3: Eigenstructure Evaporation (Lemma 2.1) - Section 4: Fixed-Point Convergence (Theorem 2.5) - Section 5: Irreversibility (Corollary 2.7) - Section 6:  $\Delta t$  Interpretation - Section 7: Related Work - Section 8: Discussion and implications

---

## 2. Formal Setup

### 2.1 State Space and Distributions

Let  $X$  be a finite set representing the system's **state space**. Elements  $x \in X$  represent possible system states, which may be:

- Content types in a recommendation system
- Output behaviors in a generative model
- Strategies in a market
- Organizational structures in an institution

We denote by  $\Delta(X)$  the **probability simplex** over  $X$ :

$$\Delta(X) = \{p: X \rightarrow [0,1] \mid \sum_x p(x) = 1\}$$

A distribution  $p \in \Delta(X)$  describes the relative prevalence of different states in the system at a given time.

### 2.2 Scalar Reward Function

A **scalar reward function** is a mapping  $r: X \rightarrow \mathbb{R}$  that assigns a real-valued reward to each state. We assume  $r$  is bounded:

$$r_{\min} \leq r(x) \leq r_{\max} \text{ for all } x \in X$$

The **reward-maximal set** is:

$$S_{\max} = \{x \in X \mid r(x) = r_{\max}\}$$

For technical convenience, we assume  $r_{\max} > r_{\min}$  (non-degenerate rewards) and  $|S_{\max}| \geq 1$  (at least one maximal state exists).

### 2.3 Multiplicative Reweighting Operator

The core of our analysis is the **multiplicative reweighting operator**  $T: \Delta(X) \rightarrow \Delta(X)$  defined by:

$$T(p)(x) = [p(x) \cdot e^{\gamma \{r(x)\}}] / Z(p)$$

where: -  $\gamma > 0$  is the **learning rate** or **temperature inverse** -  $Z(p) = \sum_x p(x) e^{\gamma \{r(x)\}}$  is the normalization constant

This operator has the following interpretation:

- States with higher reward receive exponentially amplified probability mass
- The normalization  $Z(p)$  ensures the result remains a valid probability distribution
- As  $\gamma \rightarrow \infty$ , the operator approaches a hard selection rule (winner-take-all)

**Empirical justification:** This form arises naturally in:  
- **Recommendation systems:** Softmax ranking with engagement scores  
- **Reinforcement learning:** Policy gradient updates and RLHF  
- **Evolutionary dynamics:** Replicator equations with fitness  
- **Market dynamics:** Kelly criterion and log-optimal betting

## 2.4 Closed-Loop Dynamics

We study the **closed-loop iteration**:

$$p_{\{t+1\}} = T(p_{\{t\}})$$

starting from an initial distribution  $p_0 \in \Delta(X)$ . This represents a system where:

- The current state distribution  $p_t$  generates outcomes
- Outcomes are evaluated by reward function  $r$
- The reward drives reweighting via operator  $T$
- The reweighted distribution becomes the new state

This closed loop is the mathematical formalization of “training on your own outputs,” “algorithmic feedback,” or “self-reinforcing selection.”

## 2.5 Entropy and Concentration

For a distribution  $p \in \Delta(X)$ , the **Shannon entropy** is:

$$H(p) = -\sum_x p(x) \log p(x)$$

Entropy measures the **degree of concentration** of the distribution:  
-  $H(p) = 0$  when  $p$  is a point mass (fully concentrated)  
-  $H(p) = \log |X|$  when  $p$  is uniform (maximally dispersed)

We say a distribution  $p$  is **-concentrated on  $S \subseteq X$**  if:

$$\sum_{x \in S} p(x) \geq 1 - \epsilon$$


---

## 3. Eigenstructure Evaporation

### 3.1 Statement

**Lemma 2.1 (Eigenstructure Evaporation):**

Let  $p_0 \in \Delta(X)$  be an initial distribution with  $p_0(x) > 0$  for all  $x \in X$ . Define the sequence  $\{p_t\}$  by  $p_{\{t+1\}} = T(p_{\{t\}})$ . Then for any state  $x \in S_{\text{max}}$ :

$$p_t(x) / p_t(x_{\text{max}}) = [p_0(x) / p_0(x_{\text{max}})] \cdot e^{-\gamma t (r_{\text{max}} - r(x))}$$

where  $x_{\text{max}} \in S_{\text{max}}$  is any reward-maximal state.

**Interpretation:** The relative probability of any non-maximal state decays **exponentially** with rate proportional to the reward gap  $\Delta r_x = r_{\text{max}} - r(x)$ . States with larger reward deficits decay faster. The decay is irreversible under continued application of  $T$ .

### 3.2 Proof

**Proof:**

By definition of  $T$ :

$$p_{\{t+1\}}(x) = [p_t(x) e^{\{-r(x)\}}] / Z(p_t)$$

$$\text{where } Z(p_t) = \sum_y p_t(y) e^{\{-r(y)\}}.$$

Consider the ratio between a non-maximal state  $x$  and a maximal state  $x_{\max}$ :

$$p_{\{t+1\}}(x) / p_{\{t+1\}}(x_{\max}) = [p_t(x) e^{\{-r(x)\}}] / [p_t(x_{\max}) e^{\{-r_{\max}\}}]$$

Simplifying:

$$p_{\{t+1\}}(x) / p_{\{t+1\}}(x_{\max}) = [p_t(x) / p_t(x_{\max})] \cdot e^{\{-[r(x) - r_{\max}]\}}$$

Since  $r(x) < r_{\max}$  for  $x \in S_{\max}$ , we have  $r(x) - r_{\max} = -\Delta r_x < 0$ .

Define the ratio sequence  $R_t(x) = p_t(x) / p_t(x_{\max})$ . Then:

$$R_{\{t+1\}}(x) = R_t(x) \cdot e^{\{-\Delta r_x\}}$$

This is a geometric sequence with ratio  $\gamma_x = e^{\{-\Delta r_x\}} < 1$ . Unrolling:

$$R_t(x) = R_0(x) \cdot \gamma_x^t = R_0(x) \cdot e^{\{-t \Delta r_x\}}$$

Substituting back:

$$p_t(x) / p_t(x_{\max}) = [p_0(x) / p_0(x_{\max})] \cdot e^{\{-t(r_{\max} - r(x))\}}$$

This completes the proof.

### 3.3 Entropy Corollary

**Corollary 3.1 (Entropy Reduction):**

Under the dynamics  $p_{\{t+1\}} = T(p_t)$ , the Shannon entropy  $H(p_t)$  is nonincreasing. If the initial distribution  $p_0$  has positive mass outside  $S_{\max}$ , then  $H(p_{\infty}) < H(p_0)$ , where  $p_{\infty}$  is the limit distribution.

**Proof sketch:**

The multiplicative reweighting operator  $T$  transfers probability mass from low-reward states to high-reward states. By Lemma 2.1, mass outside  $S_{\max}$  decays exponentially, concentrating the distribution. Since entropy is maximized when probability is uniformly distributed and decreases under concentration, entropy is nonincreasing. Once all mass is concentrated on  $S_{\max}$ , the operator  $T$  acts as the identity on distributions supported only on  $S_{\max}$  (since all states in  $S_{\max}$  have equal reward  $r_{\max}$ ), so entropy stabilizes. The final entropy  $H(p_{\infty}) = \log |S_{\max}|$  is strictly lower than  $H(p_0)$  whenever  $p_0$  had support outside  $S_{\max}$ .

## 4. Fixed-Point Convergence

### 4.1 Statement

**Theorem 2.5 (Fixed-Point Convergence):**

Let  $p_0 \in \Delta(X)$  with  $p_0(x) > 0$  for all  $x \in X$ . The sequence  $\{p_t\}$  defined by  $p_{t+1} = T(p_t)$  converges to a fixed point  $p_\infty \in \Delta(X)$  characterized by:

1. **Concentration:**  $\text{supp}(p_\infty) \subseteq S_{\max}$
2. **Entropy reduction:** If  $p_0$  has positive mass outside  $S_{\max}$ , then  $H(p_\infty) < H(p_0)$
3. **Exponential convergence to  $S_{\max}$ :** The mass outside  $S_{\max}$  decays as:

$$\sum_{x \notin S_{\max}} p_t(x) \leq C \cdot e^{-t \Delta r_{\min}}$$

where  $\Delta r_{\min} = \min_{x \notin S_{\max}} (r_{\max} - r(x))$  and  $C$  depends on  $p_0$ .

**Interpretation:** The system collapses onto the reward-maximal set  $S_{\max}$ , eliminating all non-maximal modes. The exact distribution over  $S_{\max}$  depends on initial conditions, but all fixed points have support contained in  $S_{\max}$  and lower entropy than any initial distribution with mass outside  $S_{\max}$ .

### 4.2 Proof

**Proof:**

**Step 1: Fixed points must be supported on  $S_{\max}$ .**

Suppose  $p^*$  is a fixed point. Then  $T(p) = p$ , which means:

$$p(x) = [p^*(x) e^{\{-r(x)\}}] / Z(p)$$

For any  $x$  with  $p^*(x) > 0$ , rearranging gives:

$$Z(p^*) = e^{\{-r(x)\}}$$

Since  $Z(p)$  is a single value, all states in  $\text{supp}(p)$  must have the same reward value. The only consistent value is  $r_{\max}$ , since if any state  $x$  with  $r(x) < r_{\max}$  had  $p(x) > 0$ , its ratio relative to states with reward  $r_{\max}$  would decay under  $T$ , contradicting the fixed-point condition. Therefore  $\text{supp}(p) \subseteq S_{\max}$ .

**Step 2: Any distribution on  $S_{\max}$  is a fixed point.**

For states  $x \in S_{\max}$ , all have  $r(x) = r_{\max}$ . Therefore, for any distribution  $p$  supported on  $S_{\max}$ :

$$\begin{aligned} T(p)(x) &= [p(x) e^{\{-r_{\max}\}}] / [\sum_{y \in S_{\max}} p(y) e^{\{-r_{\max}\}}] = [p(x) e^{\{-r_{\max}\}}] / [e^{\{-r_{\max}\}} \sum_{y \in S_{\max}} p(y)] = p(x) / \sum_{y \in S_{\max}} p(y) \\ &= p(x) \end{aligned}$$

Thus  $T$  acts as the identity operator on the face of the simplex corresponding to distributions over  $S_{\max}$ . Any distribution supported entirely on  $S_{\max}$  is a fixed point. In particular, if  $p$  assigns zero mass to some  $x \in S_{\max}$ , then  $T(p)(x) = 0 \cdot e^{\{-r_{\max}\}} = 0$ , so distributions supported on any subset of  $S_{\max}$  are also fixed points. The specific fixed point reached depends on the initial distribution  $p_0$  and the trajectory of convergence.

### Step 3: Convergence to the $S_{\max}$ face.

From Lemma 2.1, for any  $x \in S_{\max}$ :

$$p_t(x) / p_t(x_{\max}) = [p_0(x) / p_0(x_{\max})] \cdot e^{\{-t(r_{\max} - r(x))\}}$$

Since  $p_t(x_{\max})$  is bounded below (as the system concentrates mass on  $S_{\max}$ , the total mass on  $S_{\max}$  approaches 1), we have:

$$p_t(x) \leq C_x \cdot e^{\{-t\Delta r_x\}}$$

for some constant  $C_x$  depending on  $p_0$ .

The total mass on non-maximal states is:

$$\sum_{x \in S_{\max}} p_t(x) = \sum_{x \in S_{\max}} C_x \cdot e^{\{-t\Delta r_x\}} \leq C \cdot e^{\{-t\Delta r_{\min}\}}$$

where  $C = \sum_x C_x$  and  $\Delta r_{\min} = \min_{x \in S_{\max}} \Delta r_x$ .

Since total mass is conserved, the mass on  $S_{\max}$  satisfies:

$$\sum_{x \in S_{\max}} p_t(x) = 1 - C \cdot e^{\{-t\Delta r_{\min}\}}$$

Therefore the distribution converges exponentially fast to the face of the simplex corresponding to distributions over  $S_{\max}$ .

### Step 4: Entropy reduction.

By Corollary 3.1,  $H(p_t)$  is nonincreasing and  $H(p_{\infty}) < H(p_0)$  whenever  $p_0$  has positive mass outside  $S_{\max}$ . Since any distribution supported on  $S_{\max}$  satisfies  $H(p_{\infty}) = \log |S_{\max}|$ , this completes the characterization.

This completes the proof.

## 4.3 Interpretation

**Theorem 2.5 tells us that closed-loop scalar optimization inevitably produces:**

1. **Mode suppression:** All non-reward-maximal states are eliminated from the support
2. **Entropy reduction:** The distribution becomes more concentrated, losing diversity
3. **Exponential timescale:** The collapse occurs at rate proportional to  $\Delta r_{\min}$

4. **Face collapse:** The system collapses to a lower-dimensional face of the probability simplex

**This is not a failure mode—it is the designed behavior of the operator  $T$ .** Any system using multiplicative reweighting by scalar reward will exhibit this collapse, regardless of domain or implementation details. The exact distribution on  $S_{\text{max}}$  is determined by the dynamics during collapse, but all trajectories lead to elimination of non-maximal modes.

---

## 5. Irreversibility

### 5.1 Statement

**Corollary 2.7 (Irreversibility):**

Let  $p_t \rightarrow p_\infty$  under the dynamics  $p_{t+1} = T(p_t)$ . Once the system is  $\epsilon$ -concentrated on  $S_{\text{max}}$  (i.e.,  $\sum_{x \in S_{\text{max}}} p_t(x) = 1 - \epsilon$  for small  $\epsilon$ ), continued application of the same operator  $T$  cannot restore suppressed modes. Specifically:

1. **No spontaneous recovery:** For  $x \in S_{\text{max}}$  with  $p_t(x) > 0$ , we have  $p_{t+k}(x) = e^{-k \Delta r_x}$  for all  $k \geq 0$ . The mode can only decay further.
2. **Structural requirement for recovery:** To restore  $p(x)$  for  $x \in S_{\text{max}}$  requires one of:
  - **Exogenous forcing:** Direct injection of mass into  $x$  from outside the system
  - **Reward modification:** Changing  $r$  such that  $x$  enters  $S_{\text{max}}$
  - **Operator modification:** Replacing  $T$  with a different update rule
3. **Same-operator futility:** Any intervention that only adjusts parameters of  $T$  (e.g., changing  $r$ , applying  $T$  with different frequency, adding noise to  $r$ ) while keeping the same reward-maximal set  $S_{\text{max}}$  will not restore suppressed modes.

### 5.2 Proof

**Proof:**

**Part 1:** From Lemma 2.1, the ratio  $p_t(x) / p_t(x_{\text{max}})$  decays exponentially for  $x \in S_{\text{max}}$ . If  $p_t(x) > 0$  and  $p_t(x_{\text{max}}) > 0$  for some  $\epsilon > 0$  (which holds once the system is concentrated on  $S_{\text{max}}$ ), then:

$$p_{t+k}(x) / p_{t+k}(x_{\text{max}}) = [p_t(x) / p_t(x_{\text{max}})] \cdot e^{-k \Delta r_x} \\ ( / ) \cdot e^{-k \Delta r_x}$$

Since  $p_{\{t+k\}}(x_{\max})$  is bounded below by  $> 0$  (normalization ensures  $S_{\max}$  maintains mass), we have:

$$p_{\{t+k\}}(x) \propto e^{\{-k \Delta r_x\}} \cdot e^{\{-k \Delta r_x\}}$$

for appropriate choice of constants. Thus the mode continues to decay.

**Part 2:** The only way to increase  $p_t(x)$  for  $x \in S_{\max}$  is to either: - Add mass directly:  $p_t(x) \leftarrow p_t(x) + \epsilon$  for some  $\epsilon > 0$  (exogenous forcing) - Increase  $r(x)$ : make  $x$  part of the reward-maximal set - Change the operator: use a different reweighting rule

The operator  $T$  itself cannot create mass where there is none—it is a multiplicative process. Once  $p_t(x) = 0$ , we have  $T(p_t)(x) = 0$ . Furthermore, once the distribution is concentrated on  $S_{\max}$ ,  $T$  acts as the identity (since all states in  $S_{\max}$  have equal reward), preventing any change in the distribution.

**Part 3:** Adjusting  $\epsilon$  only changes the *rate* of decay, not the direction. Smaller  $\epsilon$  slows collapse but doesn't reverse it. Larger  $\epsilon$  accelerates it. Adding noise to  $r$  shifts which states are in  $S_{\max}$  but doesn't change the collapse dynamics toward the new  $S_{\max}$ . Applying  $T$  less frequently (reducing update rate) slows the collapse but doesn't prevent it.

Therefore, interventions that preserve the same  $S_{\max}$  and use the same operator  $T$  cannot restore suppressed modes.

This completes the proof.

### 5.3 Implications

**Corollary 2.7 explains why diversity interventions fail in practice:**

- “**Freshness bonuses**” in recommendation systems: Still scalar reward → same collapse dynamics
- “**Exploration phases**” in RL: Temporarily add noise, but training converges back to mode-collapsed policy
- “**Diversity metrics**” in platforms: If implemented as scalar reward adjustments, they shift  $S_{\max}$  but don't prevent collapse to the new maximum

**What actually works:** - **Human curation:** Exogenous forcing that directly injects non-maximal content - **Multi-objective optimization:** Non-scalar reward breaks the theorem's assumptions - **Structural time-gating:** Slow-mode processes that operate on different timescales than the fast optimization loop

By exogenous forcing we mean interventions that are not themselves selected, tuned, or evaluated by the scalar reward  $r$  or any functional of  $p_t$ . If a “diversity intervention” is driven by an additional scalar score and incorporated into the same optimization pipeline, it is endogenous to  $T$  and subject to the same collapse dynamics.

---

## 6. $\Delta t$ Interpretation: Temporal Collapse

### 6.1 Multi-Timescale Systems

Real systems often exhibit multiple characteristic timescales: - **Fast processes:** Immediate responses, real-time optimization, user interactions - **Slow processes:** Cultural drift, institutional evolution, expertise development

We can represent this as a distribution over timescales , where different system components evolve at different rates.

In the  $\Delta t$  framework [Beck 2024a], a system's coherence depends on maintaining **slow-mode dominance:** the ability of slow integrators to govern long-term behavior even as fast modes fluctuate.

### 6.2 Scalar Optimization as Timescale Homogenization

Scalar optimization with operator T forces all system dynamics to evolve at a single cadence: the update frequency of the optimization loop.

#### Mechanism:

- The optimization loop runs at timescale  $\Delta t_{\text{opt}}$  - States that respond at this timescale receive reward - States that evolve more slowly ( $\Delta t_{\text{slow}} \gg \Delta t_{\text{opt}}$ ) are **systematically suppressed** because they don't generate immediate reward signals

#### Result:

The system loses multi-timescale structure and collapses to a single-mode temporal regime.

### 6.3 Fixed Point as Collapsed Temporal Attractor

The fixed point  $p_\infty$  from Theorem 2.5 can be interpreted temporally:

$p_\infty$  represents a system where: - Only fast-responding, reward-maximal modes survive - Slow modes with longer integration times have been eliminated - The system has **no temporal memory** beyond the optimization timescale

This is **temporal monoculture:** the system operates exclusively at  $\Delta t_{\text{opt}}$  with no slow modes to provide resilience, adaptation, or long-horizon coherence.

### 6.4 Same- $\Delta t$ Cannot Repair

Corollary 2.7's irreversibility has a temporal interpretation:

**Operating at the same  $\Delta t$  that caused the collapse cannot restore the lost slow modes.**

Why? Because: 1. Slow modes require slow evaluation timescales ( $\Delta t_{\text{slow}}$ )  
2. Scalar optimization at  $\Delta t_{\text{opt}}$  suppresses anything not responsive at that timescale  
3. Continuing to optimize at  $\Delta t_{\text{opt}}$  cannot conjure processes that operate at  $\Delta t_{\text{slow}}$

**Recovery requires:** - Introducing new processes that operate at  $\Delta t_{\text{slow}}$  (exogenous slow integrators)  
- Changing the reward function to value slow-mode properties  
- Decoupling optimization from the fast loop (multi-timescale architecture)

## 6.5 Connection to Eigenstructure Collapse

In the  $\Delta t$  framework, **eigenstructure collapse** refers to the loss of slow eigenmodes in a dynamical system's spectral decomposition.

The scalar reward collapse theorems provide the **mechanism** for this collapse: - Scalar optimization via  $T$  is a contraction that eliminates all eigenmodes except those aligned with  $S_{\max}$  - The surviving eigenmode corresponds to the fast optimization loop - Slow eigenmodes (representing long-timescale processes) are exponentially suppressed

**Theorem 2.5 is the formal proof that eigenstructure collapse is inevitable under scalar optimization in closed loops.**

---

## 7. Related Work

### 7.1 Multiplicative Weights and Online Learning

The operator  $T$  studied in this paper belongs to the well-known family of **multiplicative weights update** (MWU) algorithms [Arora et al. 2012, Freund & Schapire 1997]. In online learning theory, MWU methods are analyzed for their regret bounds and convergence to equilibria in games. Classic results show that such algorithms converge to Nash equilibria in zero-sum games and best-response dynamics in general games. The mathematical structure has deep roots: Bush & Mosteller (1953) developed stochastic learning models with this form for psychology experiments, predating modern learning theory by decades.

**Our contribution:** While the convergence properties of MWU are well-studied in the context of optimization performance, we reframe the operator as a **diversity-destroying mechanism** in closed-loop settings and prove irreversibility results that are not standard in the learning theory literature. The focus shifts from “how fast does it converge to good strategies” to “what structural damage does it cause to the state space.”

## 7.2 Replicator Dynamics and Evolutionary Game Theory

The dynamics  $p_{\{t+1\}} = T(p_{\{t\}})$  are formally equivalent to the **discrete-time replicator equation** from evolutionary game theory [Hofbauer & Sigmund 1998, Weibull 1995]. In that context,  $r(x)$  represents fitness and  $T$  describes frequency-dependent selection. Classical results characterize evolutionarily stable strategies (ESS) and convergence to fitness peaks.

**Our contribution:** Evolutionary game theory typically assumes static or frequency-dependent fitness landscapes and studies equilibrium selection. We extend this framework to **closed-loop systems** where the reward function  $r$  is fixed but the system “trains on its own outputs,” producing what we term autophagic feedback. The irreversibility theorem (Corollary 2.7) formalizes why genetic diversity loss under strong selection is permanent—a result stated informally in evolutionary biology but not, to our knowledge, proven as a general dynamical theorem with explicit recovery conditions.

## 7.3 Mode Collapse in Machine Learning

Mode collapse is a recognized pathology in generative adversarial networks (GANs) [Goodfellow et al. 2014] and has been observed in reinforcement learning with human feedback (RLHF) [Ouyang et al. 2022]. Existing work treats mode collapse as a failure mode of specific architectures or training procedures, with proposed solutions including regularization, diversity penalties, and multi-objective optimization.

**Our contribution:** We prove that mode collapse is **not architecture-specific** but rather a general property of scalar reward optimization in closed loops. Theorem 2.5 establishes that any system using multiplicative reweighting by a scalar reward will exhibit mode suppression and convergence to concentrated distributions. This explains why ad-hoc fixes (entropy bonuses, exploration phases) provide only temporary relief—they do not change the fundamental operator structure. The irreversibility result further shows why “better tuning” cannot restore lost modes.

## 7.4 Information Theory and Entropy Minimization

Our entropy reduction result (Corollary 3.1) connects to the broader literature on entropy minimization under constraints [Cover & Thomas 2006]. In statistical mechanics, the Boltzmann distribution minimizes free energy subject to energy constraints; our operator  $T$  can be viewed as the zero-temperature limit of this process.

**Our contribution:** We apply entropy minimization analysis to **closed-loop social and algorithmic systems** rather than physical systems, and prove that the resulting entropy-reduced fixed points represent informational collapse rather than thermodynamic equilibrium. The  $\Delta t$  interpretation (Section 6)

further shows that entropy reduction in the state space corresponds to timescale homogenization—a connection not present in classical statistical mechanics.

### 7.5 Mirror Descent and Exponential Weights

$T$  is a special case of **mirror descent** [Beck & Teboulle 2003] with the negative entropy (KL divergence) as the Bregman divergence. Mirror descent algorithms are widely used in optimization and have well-studied convergence properties.

**Our contribution:** Standard mirror descent analysis focuses on convergence to optima and regret bounds. We instead analyze the **destructive side effects** of applying mirror descent in closed loops: the systematic elimination of non-optimal modes and the irreversibility of this elimination. Our results show that mirror descent is not merely an optimization algorithm but an **entropy-reducing attractor** that permanently reshapes the state space.

### 7.6 Positioning and Novel Claims

To summarize, while the operator  $T$  and its convergence properties are well-known in various forms across multiple literatures, our work makes the following novel contributions:

1. **Reframing scalar optimization as a collapse operator** rather than a learning optimizer
2. **Proving irreversibility** (Corollary 2.7) as a formal theorem with explicit recovery conditions
3. **Connecting collapse to timescale structure** via the  $\Delta t$  framework
4. **Applying the results across domains** (platforms, AI, markets, institutions) to show that diverse pathologies share a common mathematical origin
5. **Providing a negative result:** scalar optimization in closed loops cannot maintain diversity, regardless of tuning or regularization

The theorems establish that observed pathologies including “mode collapse,” “platform homogenization,” and “institutional monoculture” are not domain-specific failures but **mathematically necessary outcomes** of a ubiquitous operator structure.

---

## 8. Discussion

### 8.1 Generality and Scope

The theorems proven here apply to any system with: 1. A finite state space  $X$  2. A scalar reward function  $r: X \rightarrow \mathbb{R}$  3. A multiplicative reweighting operator  $T$  4. Closed-loop dynamics  $p_{t+1} = T(p_t)$

This structure appears in:

- **Recommendation systems** (content types, engagement scores, ranking algorithms)
- **Reinforcement learning** (action spaces, reward functions, policy updates)
- **Evolutionary dynamics** (genotypes, fitness, selection)
- **Market dynamics** (strategies, payoffs, imitation)
- **Organizational optimization** (structures, performance metrics, selection pressure)

The theorems are **domain-agnostic**. The mathematics does not care whether  $X$  represents content, behaviors, strategies, or structures. The collapse is determined entirely by the operator structure.

Real systems can temporarily delay collapse through exogenous forcing (e.g., human curation, novelty injection) or density-dependent rewards. These mechanisms act as active counterweights to the intrinsic dynamics proven here. The theorems characterize the baseline: what the system converges to when those counterweights weaken, fail, or are removed. This provides a principled benchmark for evaluating whether an intervention materially alters the operator, or merely slows the underlying collapse dynamics. Modern platform optimization increasingly automates, cost-cuts, or eliminates such counterweights in favor of algorithmic efficiency, accelerating the convergence to the collapsed fixed point.

## 8.2 Continuous State Spaces

While we focused on finite  $X$  for technical simplicity, the results extend to continuous state spaces with appropriate measure-theoretic formulations:

- Replace  $\Delta(X)$  with probability measures on a manifold
- Replace sums with integrals
- $T$  becomes a measure-valued operator
- The key inequality (eigenstructure evaporation) holds in Radon-Nikodym derivative form

The qualitative conclusions remain unchanged: scalar optimization produces mode suppression, fixed-point convergence to concentrated distributions, and irreversibility.

## 8.3 Time-Varying Rewards

If the reward function changes over time,  $r_t(x)$ , the system dynamics become:

$$p_{t+1} = T_t(p_t) \text{ where } T_t \text{ uses } r_t$$

The theorems still apply *locally*: if  $r_t$  is slowly varying relative to the optimization timescale, the system tracks a quasi-equilibrium that collapses onto  $S_{\max}(t)$ .

If  $r_t$  changes rapidly (comparable to or faster than the optimization rate), the system may never fully collapse but will exhibit **chaotic mode-switching** between different  $S_{\max}(t)$ . This is arguably worse—no stable attractor, just turbulent concentration on whichever states currently maximize reward.

## 8.4 Relationship to Other Results

### Connection to thermodynamics:

The entropy decrease guaranteed by Corollary 3.1 mirrors the second law of thermodynamics, but operates in information space rather than physical space. Scalar optimization is an *anti-thermodynamic* process: it creates order (concentration) at the expense of diversity.

### Connection to learning theory:

In machine learning, mode collapse is a known failure mode of GANs and other generative models. Our theorems show that mode collapse is not specific to GANs—it is a *general property of scalar optimization*, and RLHF is not exempt.

### Connection to evolutionary dynamics:

The replicator equation in evolutionary game theory has the same form as our operator  $T$ . Our theorems formalize why evolution under a single fitness criterion leads to loss of genetic diversity (genetic drift toward fitness peaks).

### Connection to portfolio theory:

The Kelly criterion and log-optimal betting produce multiplicative wealth reweighting. Our theorems explain why markets dominated by algorithmic traders converge to herding behavior—they are running  $T$  on strategy space.

### Remark (LLM Style Collapse).

In the special case where  $X$  is the model’s token-level continuation space and  $r$  is a scalar human preference signal, repeated application of  $T$  induces convergence toward the minimal-entropy continuation distribution. This phenomenon, known empirically as “beige prose,” is an instance of Theorem 2.5.

## 8.5 Limitations

Our analysis makes several simplifying assumptions:

**Finite state space:** Real systems may have continuous or high-dimensional state spaces. The finite case provides intuition; measure-theoretic extensions address continuous settings.

**Fixed :** We assume constant learning rate. Variable  $(t)$  can be analyzed but introduces additional complexity.

**No delays:** We assume instantaneous reward feedback. Delayed rewards ( $r_{-t}$  depends on  $p_{\{t-k\}}$ ) create memory effects that can slow but not prevent collapse.

**No noise:** We assume deterministic dynamics. Stochastic perturbations can widen the collapsed attractor slightly but don’t fundamentally change the picture.

**Single reward:** The key assumption. Multi-objective optimization with **vector rewards** breaks the theorem’s applicability—this is the primary escape route from collapse.

### **Multi-Objective Escape Route.**

The collapse theorems rely critically on scalar reward. If  $r: X \rightarrow \mathbb{R}^n$  is vector-valued with incommensurable objectives (i.e., the Pareto frontier has positive measure), then no scalar contraction exists that collapses the distribution onto a single reward-maximal set. Multi-objective optimization is therefore the only known structural escape from scalar collapse.

## **8.6 Implications for System Design**

The theorems provide a negative result: **scalar optimization in closed loops inevitably produces collapse**. This suggests design principles:

**What doesn't work:** - Tuning hyperparameters ( , update frequency) - Adding regularization to scalar reward - Periodic “diversity injections” without structural change

Entropy bonuses, regularization terms, and KL penalties remain scalar objectives. Adding  $H(p)$  or a KL constraint simply modifies the effective scalar reward to  $r'(x) = r(x) + \cdot g(x)$ . The operator is still scalar; collapse still occurs. These techniques can shift the identity of  $S_{\max}$  and slow convergence, but cannot prevent it.

**What might work:** - **Multi-objective optimization:** Use vector rewards  $r: X \rightarrow \mathbb{R}^n$  instead of scalar  $r: X \rightarrow \mathbb{R}$  - **Exogenous forcing:** Maintain external processes that inject non-maximal modes - **Slow-mode preservation:** Architect systems with timescale separation and protect slow integrators from fast optimization pressure - **Hybrid control:** Human-in-the-loop curation to provide non-scalar evaluation

## **8.7 Open Questions**

**Computational complexity:** How efficiently can one compute the fixed point  $p_\infty$  without simulating full dynamics?

**Transition dynamics:** Can we characterize the trajectory of  $p_t$  during collapse? Are there phase transitions or critical slowings?

**Multi-scale analysis:** How do the theorems extend to hierarchical systems with multiple optimization loops at different scales?

**Adversarial perturbations:** What perturbations to  $p_t$  maximally delay convergence to concentrated distributions on  $S_{\max}$ ?

**Reversibility under noise:** What noise regime is required to make collapse reversible or to maintain broader support?

**Quantifying exogenous forcing requirements:** A natural extension is to quantify the minimum rate of exogenous forcing required to maintain support outside  $S_{\max}$ , as a function of  $\Delta r_{\min}$ . Informally, diversity persists only while the injection rate exceeds the intrinsic collapse rate.

## 8.8 Relation to Goodhart-Type Failures

Classical formulations such as Campbell’s Law (“the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures”) and Strathern’s restatement (“when a measure becomes a target, it ceases to be a good measure”) can be read as informal descriptions of the dynamics we prove here. Eigenstructure Evaporation (Lemma 2.1) corresponds to the progressive corruption of a metric as probability mass shifts toward reward-maximal states; Fixed-Point Convergence (Theorem 2.5) formalizes “the measure becomes the only thing that matters” as collapse of the state distribution onto the reward-maximal face; and Irreversibility (Corollary 2.7) captures the intuition that such corruption cannot be undone by further use of the same metric.

---

## 9. Conclusion

We have established three fundamental theorems governing scalar reward optimization in closed loops:

1. **Eigenstructure Evaporation (Lemma 2.1):** Non-maximal modes decay exponentially
2. **Fixed-Point Convergence (Theorem 2.5):** The system collapses to a distribution concentrated on  $S_{\max}$  with reduced entropy
3. **Irreversibility (Corollary 2.7):** Recovery requires structural change, not parameter tuning

These results are **mathematically necessary consequences** of the operator structure. Any system using multiplicative reweighting by scalar reward will exhibit this behavior.

The theorems provide the formal foundation for understanding observed phenomena including platform content homogenization, AI mode collapse, market herding, and institutional monoculture. They also clarify why interventions fail: **you cannot fix collapse with the same optimization architecture that caused it.**

The  $\Delta t$  interpretation (Section 6) shows that scalar optimization is fundamentally a **timescale collapse mechanism**—it eliminates slow modes and forces single-cadence temporal dynamics.

These results can be applied to diverse domains including social platforms, AI training systems, markets, and institutions, demonstrating that observed pathologies in these systems may share a common mathematical origin.

The path forward requires structural innovation: multi-objective optimization, exogenous forcing, or timescale separation. The mathematics is clear about what won’t work—tuning the scalar reward function. The challenge is build-

ing systems that can maintain diversity and adaptability despite optimization pressure.

---

## References

- Arora, S., Hazan, E., & Kale, S. (2012). “The Multiplicative Weights Update Method: a Meta-Algorithm and Applications.” *Theory of Computing*, 8(1), 121-164.
- Beck, A., & Teboulle, M. (2003). “Mirror descent and nonlinear projected subgradient methods for convex optimization.” *Operations Research Letters*, 31(3), 167-175.
- Beck, J. (2024a). “ $\Delta t$  Framework Part 1: Multi-Scale Temporal Integration in Complex Systems.” Zenodo. <https://doi.org/10.5281/zenodo.14203887>
- Beck, J. (2024b). “ $\Delta t$  Framework Part 2: Metastability and Crisis Serialization in Institutional Systems.” Zenodo. <https://doi.org/10.5281/zenodo.14210617>
- Beck, J. (2024c). “ $\Delta t$  Framework Part 3: Kinetic Theory and Phase Transitions in Information Systems.” Zenodo. <https://doi.org/10.5281/zenodo.14226991>
- Bush, R. R., & Mosteller, F. (1953). *A Stochastic Model with Applications to Learning*. Wiley.
- Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2nd ed.). Wiley-Interscience.
- Freund, Y., & Schapire, R. E. (1997). “A decision-theoretic generalization of online learning and an application to boosting.” *Journal of Computer and System Sciences*, 55(1), 119-139.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). “Generative adversarial nets.” *Advances in Neural Information Processing Systems*, 27.
- Hofbauer, J., & Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press.
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). “Training language models to follow instructions with human feedback.” *arXiv preprint arXiv:2203.02155*.
- Weibull, J. W. (1995). *Evolutionary Game Theory*. MIT Press.
-

**Acknowledgments:** This work was developed through dialogue with Claude (Anthropic) and ChatGPT (OpenAI), serving as semantic amplifiers for analytical development. All errors and interpretative choices remain the author's responsibility.