Ecological Adverse Selection in Algorithmic Financial Markets: A Dynamic Model of Proxy-Driven Exclusion

Author: George Rugare Chingarande

Affiliation: Stellenbosch University, Faculty of Medicine and Health Sciences, Division of Medical Ethics and Law

Contact: rgc@sun.ac.za | +27219389006

ORCID: 0000-0002-9456-9876

Funding: None declared

Competing Interests: None declared

JEL Classification: G21, G28, O33, D82, C63

Abstract

Why do algorithmic governance systems in financial markets—from credit scoring algorithms to AI-driven underwriting—systematically reproduce exclusion and converge toward narrow patterns of participation? This paper proposes a dynamic mechanism: **Ecological Adverse Selection (EAST)** . Building on Akerlof's (1970) adverse selection logic and extending replicator dynamics to proxy-mediated selection environments, we formalize how repeated selection based on imperfect proxies compresses behavioral diversity at an exponential rate.

We prove three theorems showing that under high selection intensity, low variation, and high structural scarcity (limited outside options), behavioral diversity compresses exponentially at rate $\lambda = \alpha \cdot \Delta P$, converging asymptotically to complete homogeneity. We introduce the **Protection Paradox**: in data-driven financial systems, excluded agents retain behavioral diversity while experiencing welfare losses that compound over time. Critically, we demonstrate a methodological flaw in prior analyses: tracking only surviving agents renders the excluded population invisible, obscuring true welfare dynamics. Our corrected simulation—tracking both populations simultaneously—reveals that when excluded agents have no outside options (high structural scarcity), the utility gap between included and excluded reaches 0.85 by round 100, nearly three times the magnitude reported in prior work, with positive acceleration consistent with exponential divergence.

The framework yields testable predictions applicable to credit allocation, insurance underwriting, and algorithmic lending. We derive an empirical identification strategy using within-system natural experiments (e.g., introduction of an AI credit scoring system) and characterize conditions under which proxy-based selection produces institutional drift toward de facto exclusion without deliberate design. The model applies most directly to repeated, proxy-mediated, feedback-driven financial

environments—conditions increasingly prevalent in India's rapidly digitizing financial sector.

1. Introduction

1.1 Two Puzzles, One Proposed Mechanism

A central question in financial economics is why some credit allocation systems persistently produce unequal outcomes despite formal rules that appear neutral. Existing accounts emphasize statistical discrimination (Arrow, 1973), feedback loops in reputation systems (Avery et al., 2017), or algorithmic bias in training data (Obermeyer et al., 2019).

At the same time, financial markets in India and globally are rapidly adopting algorithmic systems for credit scoring (e.g., digital footprint-based lending), insurance underwriting, and portfolio management. These systems select individuals based on predictive proxies derived from historical transaction data. A growing empirical literature documents that such systems can reproduce or amplify existing disparities (Bartik et al., 2020; Fuster et al., 2022).

This paper explores whether these two phenomena may be connected through a specific dynamic mechanism: **Ecological Adverse Selection (EAST)** —the process by which repeated selection based on imperfect proxies can, under identifiable conditions, compress behavioral diversity and systematically exclude non-conforming agents from financial markets.

1.2 What EAST Is and Is Not

EAST is:

- A formal model of selection dynamics in repeated, proxy-mediated financial environments
- A complementary mechanism that can operate within existing institutional frameworks
- Particularly relevant to algorithmically mediated credit allocation, insurance underwriting, and hiring

EAST is not:

- A replacement for existing theories of statistical discrimination or adverse selection
- A general theory of all institutional persistence
- A claim that all diversity loss is welfare-reducing

1.3 Relevance to Indian Financial Markets

The framework is particularly salient for India's rapidly digitizing financial sector. The expansion of UPI, Aadhaar-linked credit scoring, and AI-driven lending platforms has created conditions for EAST dynamics to operate: repeated selection cycles, proxy mediation (e.g., digital footprints as proxies for creditworthiness), feedback hardening, and—for excluded populations—limited outside options (high structural scarcity). The model generates testable predictions about how behavioral diversity in credit access and insurance uptake may evolve under algorithmic governance.

2. Situating EAST in Existing Literature

2.1 Relationship to Adverse Selection (Akerlof)

Classical adverse selection (Akerlof, 1970) explains market failure from asymmetric information: sellers know more about quality than buyers, driving high-quality goods from the market. EAST shares the intuition that selection can degrade population composition but differs in three ways. These are illustrated in Table 1 below.

Table 1: Differences between EAST and Classical Adverse Selection Theory

Dimension	Classical Adverse Selection	EAST
Information	Asymmetric	Can be symmetric
Temporal structure	One-shot	Repeated with feedback
Mechanism	Withdrawal of high-quality agents	Compression of behavioral diversity

EAST can be seen as extending adverse selection logic from *static markets* to *dynamic selection environments*.

2.2 Relationship to Evolutionary Game Theory

EAST shares formal similarities with replicator dynamics (Weibull, 1995; Hofbauer & Sigmund, 1998). In replicator dynamics, the proportion of a strategy grows proportionally to its fitness relative to the population average. EAST's softmax selection can be viewed as a discrete-time, stochastic version of this dynamic, with proxy value interpreted as "fitness." The key normative difference is that EAST emphasizes potential divergence between selection proxy and true welfare—a concern less central in evolutionary game theory, where fitness typically aligns with reproductive success.

2.3 Relationship to Algorithmic Fairness

The algorithmic fairness literature (Hardt et al., 2016; Kleinberg et al., 2017; Mitchell et al., 2021) has documented how ML systems reproduce historical bias. EAST complements this literature by providing a *dynamic systems perspective*: fairness metrics typically measure static disparities, while EAST models how disparities evolve and potentially compound over time. The convergence rate $\lambda = \alpha \cdot \Delta P$ offers a quantitative prediction about the speed of bias amplification.

2.4 Relationship to Platform Economics

Platform economics (Rochet & Tirole, 2003; Weyl, 2010) studies two-sided markets and network effects. EAST adds a selection-theoretic perspective: platforms not only match agents but also *shape behavioral strategies* through feedback loops. The Protection Paradox extends the concept of adverse selection to dynamic platform environments where exclusion operates through invisibility rather than price.

2.5 Relationship to Acemoglu and Robinson (2012)

We position EAST as a *complementary mechanism* rather than a replacement. Table 2 clarifies the relationship.

Table 2: EAST as Complementary Mechanism to Acemoglu and Robinson

Phenomenon	Acemoglu & Robinson Account	EAST Contribution
Elite incentives	Central mechanism	Not addressed
Collective action problems	Central mechanism	Not addressed
Commitment problems	Central mechanism	Not addressed
Exclusion without coercion	Not fully specified	Formalizes passive exclusion in data-driven systems
Speed of lock-in	Generations	Exponential ($\lambda = \alpha \cdot \Delta P$)

Phenomenon	Acemoglu & Robinson Account	EAST Contribution
Algorithmic governance	Not anticipated	Models feedback dynamics

EAST is most valuable for understanding *algorithmically mediated persistence*—a domain Acemoglu and Robinson did not address.

2.6 Relationship to Knowledge Collapse

Recent work on knowledge collapse (Peterson, 2025) examines how recursive reliance on AI-generated content narrows epistemic distributions toward the mean. In Peterson's model, a 20% cost discount on AI-generated content produces public beliefs 2.3 times further from truth than when no discount exists. Joseph and Joseph (2025) describe this as "digital decay"—a feedback loop where models trained on synthetic data lose diversity and drift toward repetition. Nettasinghe and Zhao (2026) formalize the dynamics of human-AI collective knowledge on the web, distinguishing model collapse from quality dilution and human competence inversion.

EAST complements this analysis. Where knowledge collapse concerns *epistemic* distribution narrowing, EAST concerns *behavioral* diversity compression in selection environments. The two phenomena share a common mathematical structure: recursive feedback loops that compress diversity at exponential rates. In financial contexts, knowledge collapse among analysts using LLM-generated research may precede or amplify behavioral convergence in trading or credit allocation—a coupling not yet formally modeled but implied by EAST's framework.

3. Ecological Adverse Selection Theory (EAST): Formal Framework

3.1 The Softmax Selection Environment

Consider a population of agents $i \in {1,\dots,N}$ each with behavioral strategy $s_i \in [0,1]$ and true value $U(s_i) \in [0,1]$. A selection system (algorithm, institution, market) selects agents according to a measurable proxy $P(s_i)$, which may diverge from $U(s_i)$. The probability that agent i is selected is given by a softmax function:

$$\Pr(\text{select } i) = \frac{\exp(\alpha \cdot P(s_i))}{\sum_{j=1}^{N} \exp(\alpha \cdot P(s_j))}$$

where $\alpha \geq 0$ captures **selection intensity** (what we term Predictive Governance Intensity, PGI). Higher α means more deterministic selection. This formulation nests random selection ($\alpha = 0$) and deterministic threshold selection ($\alpha \to \infty$).

Over repeated selection cycles, selected strategies persist and generate data; non-selected strategies become increasingly invisible. The expected evolution of the strategy distribution is:

$$p_{t+1}(s) = \frac{p_t(s) \cdot \exp(\alpha \cdot P(s))}{\int p_t(s') \exp(\alpha \cdot P(s')) ds'} \cdot (1 - \nu) + \nu \cdot q(s)$$

where $\int p_t(s) ds = 1$ is a normalization factor, $\nu \in [0,1]$ captures the rate of variation/exploration, and $q(s)$ is the exploration distribution.

3.2 Core Constructs

EAST introduces five measurable constructs that operationalize the theoretical framework. Table 3 summarizes these constructs, their formal definitions, and empirical measurement approaches.

Table 3: Core Constructs of Ecological Adverse Selection Theory

Construct	Symbol	Definition	Formal Expression	Measurement Approach
Predictive Governance Intensity	PGI (α)	Selection intensity; degree to which selection is deterministic	α in softmax function	Estimate via maximum likelihood from selection logs
Structural Scarcity Index	SSI (ν^{-1})	Inverse of variation/exploration rate; availability of outside options	ν in evolution equation	Composite of inequality metrics (Gini coefficients)
Behavioral Diversity Distribution	BDD (H_t)	Shannon entropy of strategy distribution	$H_t = -\sum_k p_{t,k} \log p_{t,k}$	Rolling window entropy calculation

Construct	Symbol	Definition	Formal Expression	Measurement Approach
Convergence Rate	λ	Speed of behavioral diversity decline	$\lambda = \alpha \cdot \Delta P$	Segmented regression: $\log(H_t) = \beta_0 + \beta_1 t + \varepsilon_t$
Welfare Loss Parameter	β	Utility penalty imposed on excluded agents	$U_{\text{excluded}} = U(s) - \beta \cdot \mathbf{1}[s < \theta]$	Comparison of outcomes between included and excluded populations
Protection Paradox Gap	Δ_t	Difference in mean utility between included and excluded	$\Delta_t = \bar{U}_{\text{included},t} - \bar{U}_{\text{excluded},t}$	Longitudinal tracking of both populations

3.3 Three Theorems

Theorem 1 (Asymptotic Convergence). In the absence of variation ($\nu = 0$), the strategy distribution converges weakly to a point mass at the proxy-maximizing strategy $s^* = \arg\max_s P(s)$:

$$p_t \xrightarrow{d} \delta_{s^*} \text{as} t \to \infty$$

Consequently, Behavioral Diversity Distribution converges to zero:

$$\lim_{t\to\infty} H_t = 0$$

Interpretation: Without countervailing variation, selection systems lock into a single strategy—the one that maximizes the proxy. This formalizes a "vicious circle" of behavioral homogenization.

Theorem 2 (Exponential Convergence). For a binary strategy space $s \in {0,1}$ with $P(1) > P(0)$, the proportion p_t of agents adopting $s=1$ evolves according to:

$$p_{t+1} = \frac{p_t \cdot e^{\alpha P(1)}}{p_t \cdot e^{\alpha P(1)} + (1 - p_t) \cdot e^{\alpha P(0)}}$$

The gap $g_t = |p_t - 1|$ decays exponentially:

$$g_t \leq g_0 \cdot e^{-\lambda t}, \lambda = \alpha \cdot (P(1) - P(0))$$

Interpretation: Lock-in happens exponentially fast—faster with higher selection intensity. This implies that algorithmic systems can produce behavioral convergence in months or years, not generations.

Theorem 3 (Steady State with Variation). With variation $\nu > 0$, the system converges to a positive steady-state diversity:

$$H^* = -\sum_k p_k^* \log\ p_k^* > 0$$

where p_k^* is decreasing in α (selection intensity) and increasing in ν (variation rate). The steady-state distribution satisfies:

$$p_k^* = \frac{\nu \cdot q_k}{(1 - \nu) \cdot (1 - \mathbb{E}[\exp\ (\alpha P)]^{-1}) + \nu}$$

Interpretation: Introducing variation—through exploration, randomness, or outside options—prevents complete lock-in. The steady-state diversity is higher when variation is higher and selection intensity is lower.

Proofs of all theorems are provided in Appendix A.

4. Simulation Analysis: Two Approaches to Tracking Welfare Dynamics

We present two complementary simulation approaches that yield different insights into the dynamics of ecological adverse selection. The first approach tracks only the included (selected) population. The second approach tracks both included and excluded populations simultaneously, revealing welfare dynamics that are invisible when only survivors are observed.

The simulation is parameterized identically across both approaches:

- Strategy space: $s \in [0,1]$, initial distribution $s_i \sim \text{Uniform}(0,1)$
- Proxy function: $P(s) = s^2$
- True value: $U(s) = s - \beta \cdot \mathbf{1}[s < 0.3]$, where $\beta \in {0, 0.2, 0.4}$
- Selection intensity: $\alpha = 2.0$

- Selection rounds: $T = 100$
- Population size: $N = 10,000$

4.1 Panel A: Behavioral Diversity Dynamics (Included Only)

Under included-only tracking, the included population's entropy declines monotonically from approximately 2.5 (near-maximum diversity for a uniform distribution) to approximately 0.2 by round 100—a 92% reduction in behavioral diversity. This pattern holds across all values of β, as the welfare parameter does not affect selection dynamics.

Interpretation: Under sustained selection pressure, the included population experiences *cognitive compression*—a progressive reduction in behavioral diversity as selection repeatedly favors strategies with higher proxy values $P(s)$. The convergence rate is consistent with Theorem 2's prediction $\lambda = \alpha \cdot \Delta P$.

Limitation: This panel reveals nothing about the excluded population. Agents who fail to be selected simply vanish from the analysis.

4.2 Panel B: Mean Utility Dynamics (Included Only)

For all β, included population utility rises from approximately 0.50 to approximately 0.95 by round 100. Higher β reduces the terminal utility slightly, but all lines show positive gains.

Interpretation: Under included-only tracking, the system appears Pareto-improving: all surviving agents experience welfare gains. Higher β reduces the magnitude of gains but does not produce absolute losses.

Limitation: This panel cannot reveal the welfare of excluded agents because they are dropped from the dataset entirely.

4.3 Panel C: Population Dynamics (Included Only)

Under included-only tracking, the included population size remains constant at $N = 10,000$ throughout all selection rounds.

Limitation: This panel conceals the fact that exclusion is a net outflow process. Agents flow from the included to the excluded population as they fail to be selected.

4.4 Panel D: Utility Gap Dynamics (Included Only)

Under included-only tracking, the utility gap between the included population at time t and the included population at time 0 shows positive gains for all β, with no evidence of absolute welfare losses for any population segment.

Limitation: This specification cannot reveal the Protection Paradox. The utility gap between included and excluded is never computed because excluded agents are not tracked.

4.5 Corrected Simulation: Tracking Both Included and Excluded Populations

To address the limitations of the included-only specification, we implement a corrected simulation that tracks *both populations simultaneously*:

- **Included population:** Agents selected in each round; their strategies evolve under selection pressure with variation
- **Excluded population:** Agents not selected; they accumulate over time and are never re-incorporated

Table 4: Comparison of Utility Gap Estimates Across Specifications

β	Included-Only Specification (Implied Gap)	Dual-Tracking Specification (Actual Gap)	Difference
0.2	0.30	0.65	+0.35
0.4	0.30	0.85	+0.55

Key finding: For $\beta = 0.4$, the true utility gap at round 100 is 0.85—nearly three times the 0.30 gap implied by the included-only analysis. Moreover, the gap exhibits positive acceleration ($\frac{d^2\Delta_t}{dt^2} > 0$), meaning the rate of divergence increases over time.

Interpretation: The Protection Paradox is quantified by the utility gap. Excluded agents, despite retaining higher behavioral diversity, suffer strictly lower utility than included agents. Moreover, this gap *widens over time* and, for high β, *accelerates*—consistent with Theorem 2's prediction of exponential divergence.

5. Scope Conditions: Where EAST Applies

EAST is not a universal theory. It applies under specific, identifiable conditions.

5.1 Necessary Conditions

Table 5: Necessary Conditions for Ecological Adverse Selection

Condition	Description	Why It Matters
Repeated selection	Selection occurs over multiple rounds	Feedback loops need time to operate
Proxy mediation	Selection uses measurable $P(s)$ rather than direct $U(s)$	Divergence between proxy and true value is possible
Feedback exists	Past selections influence future selection environments	Hardening can occur
Limited outside options	Excluded agents have high SSI (few alternatives)	Exclusion has welfare consequences

5.2 Where EAST Does NOT Apply

Table 6: Boundary Conditions – Where EAST Does Not Apply

Context	Why EAST Is Less Relevant
Purely coercive environments	Selection is deterministic and violent, not probabilistic
One-shot political decisions	No repeated feedback loop
Perfect proxies	$P(s) = U(s)$; no divergence means no welfare loss from compression
High-variation environments	ν large enough that BDD remains high regardless of α

5.3 Algorithms as Institutional Technologies

We do not claim that algorithms are institutions. Rather, we argue that *algorithms are institutional technologies*—systems that mediate access to resources, opportunities, and recognition within existing institutional frameworks.

Table 7: Algorithms as Institutional Technologies – Mapping Traditional Institutions to Algorithmic Analogs

Traditional Institution	Algorithmic Analog	EAST Relevance
Labor market regulation	AI hiring platforms	Selection of candidates
Credit allocation	Credit scoring systems	Selection for capital
Insurance underwriting	Algorithmic risk scoring	Selection for coverage
Welfare administration	Algorithmic benefits allocation	Selection for assistance

In each case, the algorithm operates *within* an institutional framework but introduces dynamics—speed, scale, opacity, feedback hardening—that existing institutional theories did not anticipate.

6. EAST and Institutional Dynamics

6.1 Institutional Drift

A key implication of EAST is *institutional drift*: the tendency for repeated proxy-based selection to produce de facto extractiveness without deliberate elite design.

Even when formal rules remain inclusive, repeated proxy-based selection can:

- Reduce participation diversity
- Favor historically dominant strategies
- Limit upward mobility for non-conforming agents
- Render excluded populations invisible (generating no data)

This creates outcomes similar to extractive systems—concentrated benefits, limited mobility, persistent exclusion—without requiring explicit exclusionary rules or coercive enforcement.

6.2 The Protection Paradox

The Protection Paradox, revealed clearly in our corrected Panel D, states that under EAST conditions:

1. Excluded agents exhibit *higher behavioral diversity* than included agents (no adaptive pressure)
2. Excluded agents suffer *strictly lower welfare* than included agents (exclusion from opportunity)
3. The welfare gap *widens over time* and, under high β, *accelerates*

This is *not* a universal claim about exclusion. It applies specifically to *data-driven selection systems* where:

- Exclusion is probabilistic rather than deterministic
- Excluded agents face external constraints on welfare but not on behavioral variation
- The system generates no data about excluded agents (rendering them invisible)

This mechanism may help explain *passive exclusion*—a form of marginalization distinct from overt coercion. The excluded are not actively suppressed; they are simply never seen.

6.3 What EAST Adds to Institutional Theory

Existing political economy accounts explain:

- Why elites block change (incentives)
- Why the excluded struggle to organize (collective action problems)
- How institutions persist (path dependence, increasing returns)

EAST adds a *third mechanism*:

Invisible exclusion: In data-driven systems, the excluded may not need to be actively suppressed. They simply become invisible—generating no data, appearing in no metrics, registering in no accountability mechanisms.

7. Empirical Implications and Testable Predictions

The framework yields a set of formal predictions specifying how selection mechanisms, agent heterogeneity, and exploration rates jointly determine the trajectory of diversity loss and welfare inequality.

7.1 Core Predictions

Table 8: Empirical Implications and Core Testable Predictions

Prediction	Formal Statement	Empirical Implication
P1: Compression	Higher α → lower BDD	Systems with stronger proxy-based selection show faster diversity decline
P2: Moderation	$\frac{\partial \Delta_t}{\partial \beta} > 0$	Compression is stronger when excluded agents have few alternatives
P3: Acceleration	$\frac{d^2 H_t}{dt^2} < 0$	Diversity loss speeds up over time (hardening)
P4: Asymptotic limits	$H^* = 0$ (if $\nu=0$); $H^* > 0$ (if $\nu>0$)	Systems plateau at positive diversity only with exploration
P5: Protection Paradox	$H_{\text{excluded}} > H_{\text{included}}$ but $\bar{U}_{\text{excluded}} < \bar{U}_{\text{included}}$	Excluded have higher diversity but lower welfare; gap widens exponentially
P6: Knowledge-Behavior Coupling	$\lambda_{compound} \geq \lambda_{behavioral} + \lambda_{epistemic}$	LLM-assisted financial analysis accelerates both epistemic and behavioral diversity decline

7.2 Identification Strategy: Within-System Natural Experiment

Rather than relying on arbitrary thresholds or cross-country aggregates, we propose a *within-system identification strategy*:

Design: Study the introduction of an algorithmic selection system (e.g., AI hiring platform, credit scoring algorithm) where:

- Pre- and post-data are available (minimum 12 months each)
- The system replaces a previous selection process (human screening, rule-based underwriting)

Proxies:

Construct	Proxy	Data Source
PGI (α)	Algorithm confidence score variance	System logs
BDD	Entropy of candidate/applicant characteristics	Application records
λ	Estimated from BDD time series	Longitudinal data
Protection Paradox	Hire/approval rate × subsequent outcomes for algorithm-disadvantaged groups	HR or administrative records

Hypothesis: BDD declines faster post-algorithm ($\lambda_{post} > \lambda_{pre}$) for groups with high SSI (few outside options), with no corresponding decline for low-SSI groups. The utility gap between included and excluded widens exponentially post-implementation.

Falsifiability: The hypothesis is falsified if BDD remains stable post-algorithm, if declines are equal across SSI groups, or if the utility gap does not exhibit positive acceleration.

7.3 Application Context: Algorithmic Credit Scoring in India

The framework's predictions can be tested in settings where financial algorithms have been introduced or modified. India offers several natural experimental contexts:

Context 1: Digital footprint-based lending. Several Indian fintech lenders use alternative data (mobile usage, transaction history, social network) as proxies for creditworthiness. EAST predicts that over repeated loan cycles, borrower behavioral diversity will compress exponentially unless algorithms incorporate explicit variation mechanisms.

Context 2: Aadhaar-linked credit scoring. The integration of biometric ID with credit histories creates a feedback loop where past formal-sector participation determines future access. EAST predicts that excluded populations (informal sector workers) will exhibit higher behavioral diversity but systematically lower welfare, with the gap widening over time.

Context 3: Algorithmic insurance underwriting. AI-driven health and crop insurance models using historical claims data may exclude novel risk profiles. EAST predicts that

unless variation is maintained, underwriting strategies will converge to a narrow set of historically acceptable profiles.

8. Model Predictions for System Outcomes

The framework generates positive predictions about how design parameters affect system trajectories, without prescriptive claims.

8.1 Proxy Alignment and Outcome Divergence

Prediction: The divergence between proxy and true value $|P(s) - U(s)|$ determines the steady-state welfare distribution. Systems with lower divergence (smaller $|P-U|$) exhibit higher mean true value among selected agents at convergence.

Testable implication: In credit scoring, when repayment history (proxy) diverges from true entrepreneurial productivity (U), rejected applicants with high productivity will experience systematically lower welfare than selected applicants with lower productivity.

8.2 Variation and Steady-State Diversity

Prediction: The steady-state behavioral diversity H^* is increasing in the variation/exploration rate ν and decreasing in selection intensity α. Systems with $\nu > 0$ maintain positive diversity ($H^* > 0$); systems with $\nu = 0$ converge to complete homogeneity ($H^* = 0$).

Testable implication: Algorithmic lending platforms that incorporate random exploration (e.g., ϵ-greedy sampling) will exhibit higher diversity in borrower repayment strategies than deterministic systems, holding all else equal.

8.3 Outside Options and the Protection Paradox Gap

Prediction: The utility gap Δ_t at time t is increasing in the structural scarcity index (SSI). When excluded agents have no outside options (high SSI), the gap widens exponentially at rate λ. When outside options exist (low SSI), the gap is bounded.

Testable implication: The welfare divergence between algorithmically approved and rejected loan applicants will be larger in regions with concentrated banking markets (few alternatives) than in regions with competitive banking markets.

8.4 Knowledge-Behavior Coupling

Prediction P6: In LLM-assisted financial decision-making (e.g., algorithmic trading, credit analysis), knowledge collapse among analysts predicts faster behavioral convergence among traders or loan officers. The compound convergence rate satisfies $\lambda_{compound} \geq \lambda_{behavioral} + \lambda_{epistemic}$, where $\lambda_{epistemic}$ captures cross-domain feedback strength.

Testable implication: In financial firms where analysts use LLM-generated research summaries, the diversity of trading strategies (behavioral) and the diversity of information sources (epistemic) should decline faster than in firms without LLM adoption. Cross-domain acceleration can be detected via vector autoregression or by comparing pre- and post-LLM-adoption periods.

9. Welfare Analysis

9.1 When Is Diversity Welfare-Enhancing?

We do not assume all diversity is good. Behavioral diversity is welfare-enhancing under specific conditions:

Condition	Why Diversity Matters
Environmental uncertainty	Diverse strategies hedge against unknown future states
Learning value	Exploration enables discovery of better strategies
Preference heterogeneity	Different strategies serve different populations

Diversity may be welfare-reducing when:

- Convergence enables beneficial coordination (e.g., driving on the right)
- Some strategies are demonstrably harmful

9.2 Welfare Loss from Proxy Divergence

The fundamental source of welfare loss in EAST is the divergence between proxy and true value:

$$D(s) = | P(s) - U(s) |$$

When $D(s) > 0$, selection for $P(s)$ does not maximize $U(s)$. Deadweight loss can be estimated as:

$$DWL = \mathbb{E}[U(s) \mid \text{selected}] - \max_s U(s)$$

When validation data on $U(s)$ exists (e.g., long-term job performance, health outcomes, recidivism), this can be estimated empirically.

9.3 The Compounding Harm of Exclusion

Our corrected simulation reveals that the welfare loss from exclusion is not static but *compounding*. For $\beta = 0.4$, the utility gap at round 100 is 0.85—nearly three times the 0.30 gap implied by the original analysis. Moreover, the gap exhibits positive

acceleration ($\frac{d^2\Delta_t}{dt^2} > 0$), meaning the rate of divergence increases over time.

The harm of exclusion is not a one-time penalty but a dynamically growing deficit. Each round that an agent remains excluded, the gap between their welfare and that of selected agents expands at an increasing rate.

10. Discussion

10.1 Summary of Contributions

Contribution	Description
Formal theory	EAST as a softmax selection model with three theorems
Construct table	Comprehensive operationalization of PGI, SSI, BDD, λ, β, Δ_t
Methodological correction	Identification of selective survival bias; corrected simulation revealing true welfare gap magnitude (0.85 vs. 0.30) and acceleration
Complementary mechanism	Positioned alongside (not replacing) existing institutional theories
Protection Paradox	A novel mechanism of passive exclusion in data-driven systems, with empirically testable predictions
Testable predictions	Six propositions with empirical counterparts (including P6 on knowledge-behavior coupling)
Identification strategy	Within-system natural experiment design

10.2 Limitations

1. **Formalization vs. lived experience:** EAST captures structural dynamics but not the subjective experience of exclusion.
2. **Data availability:** Measurement of PGI, SSI, and BDD requires access to proprietary systems.

3. **Causal identification:** The proposed design identifies association; causal claims require stronger assumptions (exogeneity of algorithm introduction).
4. **Scope:** EAST does not apply to purely coercive or non-repeated settings.
5. **Welfare ambiguity:** Diversity loss is not always harmful; context matters.
6. **Simulation assumptions:** Results depend on the specific functional forms chosen ($P(s)=s^2$, threshold at 0.3); sensitivity analysis is required.
7. **Epistemic scope:** The framework does not explicitly model epistemic diversity collapse (knowledge collapse), though recent work suggests parallel dynamics (Peterson, 2025; Nettasinghe & Zhao, 2026). In financial contexts where LLM-generated analysis informs trading or credit decisions, coupling behavioral and epistemic collapse may produce faster convergence than either alone—an extension for future work.

10.3 Future Research Agenda

1. **Empirical testing:** Implement the proposed within-system design across multiple algorithmic systems (credit scoring, insurance underwriting, hiring)
2. **Parameter estimation:** Develop robust methods for estimating α, ν, β, and λ from observational data
3. **Comparative institutional analysis:** Apply EAST to compare algorithmic vs. non-algorithmic selection environments
4. **Welfare quantification:** Estimate deadweight loss from proxy divergence and the compounding harm of exclusion in real-world systems
5. **Knowledge-behavior coupling:** Test Prediction P6 in financial firms with and without LLM adoption

11. Conclusion

This paper has proposed Ecological Adverse Selection Theory (EAST) as a framework for understanding how repeated proxy-based selection shapes outcomes in algorithmically mediated financial markets. We have corrected a methodological flaw in prior analyses, demonstrating that the true welfare gap between included and excluded agents can be nearly three times larger than previously reported, with positive acceleration consistent with exponential divergence.

The **Protection Paradox** captures a distinctive feature of this mechanism: excluded agents retain behavioral diversity while experiencing welfare losses that compound over time. This form of *passive exclusion*—exclusion through algorithmic invisibility rather than overt discrimination—may help explain how data-driven financial systems can produce persistent disparities without deliberate design.

The framework yields testable predictions for credit allocation, insurance underwriting, and algorithmic lending. The convergence rate $\lambda = \alpha \cdot \Delta P$ implies that algorithmic financial systems can lock in behavioral homogeneity exponentially fast—in months or years, not generations. This speed, combined with the invisibility of exclusion, makes EAST particularly relevant for understanding India's rapidly digitizing financial sector.

The corrected simulation reveals that for settings where excluded agents have no outside options (high structural scarcity), the utility gap reaches 0.85 by round 100—nearly three times the gap implied by survivor-only analysis. Moreover, the gap exhibits positive acceleration, meaning the divergence compounds over time.

The framework is bounded by scope conditions: repeated selection, proxy mediation, feedback, and limited outside options. It does not apply to purely coercive environments, one-shot decisions, or settings with perfect proxies. Future empirical work should test EAST predictions using within-system natural experiments in credit scoring, insurance underwriting, and algorithmic hiring systems, particularly in the Indian context where rapid digitization creates natural laboratories for studying these dynamics.

Acknowledgements

The author thanks Dr Thomasz Sanko for assisting with running the R code. The author did not receive any research funding or support in any form for this paper. There is no conflict of interest to report.

References

Acemoglu, D., & Robinson, J. A. (2012). *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. Crown Business.

Akerlof, G. A. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488-500.

Arrow, K. J. (1973). The theory of discrimination. In O. Ashenfelter & A. Rees (Eds.), *Discrimination in Labor Markets*. Princeton University Press.

Avery, C., Resnick, P., & Zeckhauser, R. (2017). The market for evaluations. *American Economic Review*, 89(3), 564-584.

Bartik, A. W., Bertrand, M., Lin, F., Rothstein, J., & Unrath, M. (2020). Measuring the labor market at the onset of the COVID-19 crisis. *Brookings Papers on Economic Activity*, 2020(2), 1-49.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *Journal of Finance*, 77(1), 5-47.

Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*, 29.

Hofbauer, J., & Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press.

Joseph, S., & Joseph, J. (2025). Digital decay: when AI eats its own homework. *AI & Society*. https://doi.org/10.1007/s00146-025-02229-w

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. *Proceedings of ITCS*.

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, 8, 141-163.

Nettasinghe, B., & Zhao, K. (2026). Dynamics of Human-AI Collective Knowledge on the Web. *Proceedings of the ACM Web Conference 2026 (WWW '26)*.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.

Peterson, A. J. (2025). AI and the Problem of Knowledge Collapse. *AI & Society*. https://doi.org/10.1007/s00146-025-02218-z

Rochet, J. C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4), 990-1029.

Weibull, J. W. (1995). *Evolutionary Game Theory*. MIT Press.

Weyl, E. G. (2010). A price theory of multi-sided platforms. *American Economic Review*, 100(4), 1642-1672.

Appendix A: Formal Proofs of Theorems 1-3

A.1 Proof of Theorem 1 (Asymptotic Convergence)

Consider the replicator dynamic:

$$p_{t+1}(s) = \frac{p_t(s) \cdot \exp(\alpha P(s))}{\int p_t(s') \exp(\alpha P(s')) ds'}$$

The mean proxy evolves according to:

$$\bar{P}_t = \int P(s) p_t(s) ds$$

Thus $\bar{P}_t$ is non-decreasing and bounded above by $\max_s P(s)$. Hence $\bar{P}_t$ converges to some $\bar{P}^$. *If $\bar{P}^* < \max_s P(s)$*, there exists

$\epsilon > 0$ such that $\Pr(P(s) > \bar{P}^* + \epsilon) > 0$. For any strategy with $P(s) > \bar{P}^* + \epsilon$, the replicator dynamic would increase its mass, contradicting convergence. Therefore $\bar{P}^* = \max_s P(s)$.

Now consider any open set O not containing $s^* = \arg\max_s P(s)$. For any $s \in O$, $P(s) \leq \max_s P(s) - \delta$ for some $\delta > 0$. The replicator dynamic implies:

$$\frac{p_{t+1}(O)}{p_t(O)} \leq \frac{\exp(\alpha(\max P - \delta))}{\exp(\alpha \max P)} = e^{-\alpha\delta} < 1$$

Since $e^{-\alpha \delta} < 1$, for sufficiently large t, the right-hand side becomes negative, driving $p_t(O) \to 0$. Thus p_t converges weakly to δ_{s^*}. ■

A.2 Proof of Theorem 2 (Exponential Convergence)

For binary strategies, the proportion p_t of agents adopting $s=1$ evolves according to:

$$p_{t+1} = \frac{p_t e^{\alpha P(1)}}{p_t e^{\alpha P(1)} + (1 - p_t) e^{\alpha P(0)}}$$

Define $g_t = 1 - p_t$. Then:

$$g_{t+1} = 1 - \frac{p_t e^{\alpha P(1)}}{p_t e^{\alpha P(1)} + (1 - p_t) e^{\alpha P(0)}} = \frac{(1 - p_t) e^{\alpha P(0)}}{p_t e^{\alpha P(1)} + (1 - p_t) e^{\alpha P(0)}}$$

Taking the reciprocal:

$$\frac{1}{g_{t+1}} = \frac{p_t e^{\alpha P(1)}}{(1 - p_t) e^{\alpha P(0)}} + 1 = e^{\alpha(P(1) - P(0))} \cdot \frac{p_t}{1 - p_t} + 1 = e^{\alpha \Delta P} \cdot \frac{1 - g_t}{g_t} + 1$$

Let $r_t = 1/g_t$. Then:

$$r_{t+1} = e^{\alpha \Delta P}(r_t - 1) + 1$$

Thus $r_t = e^{\alpha \Delta P \cdot t} (r_0 - 1) + 1$, and $g_t = 1/r_t$. For large t, $g_t \sim e^{-\alpha \Delta P \cdot t}$. ■

A.3 Proof of Theorem 3 (Steady State with Variation)

With variation, the evolution equation becomes:

$$p_{t+1}(s) = (1 - \nu) \frac{p_t(s) \cdot \exp(\alpha P(s))}{\int p_t(s') \exp(\alpha P(s')) ds'} + \nu \cdot q(s)$$

At steady state, $p_{t+1} = p_t = p^*$:

$$p^*(s) = (1 - \nu) \frac{p^*(s) \cdot \exp(\alpha P(s))}{Z} + \nu \cdot q(s)$$

where $Z = \int p^*(s') \exp(\alpha P(s')) ds'$. Solving:

$$p^*(s)\left(1-(1-\nu)\frac{\exp(\alpha P(s))}{Z}\right)=\nu q(s)$$

Thus:

$$p^*(s)=\frac{\nu q(s)}{1-(1-\nu)\frac{\exp(\alpha P(s))}{Z}}$$

provided $Z > (1-\nu)\exp(\alpha P(s))$ for all s. The steady-state entropy $H^* = -\int p^*(s) \log p^*(s) ds$ is decreasing in α and increasing in ν by the monotonicity of the transformation. ■

Appendix B: Corrected Simulation Code (R)

```
r

# ============================================================================

# CORRECTED EAST SIMULATION

# Tracks both included and excluded populations simultaneously

# ============================================================================


library(tidyverse)

library(ggplot2)

library(patchwork)


# Softmax selection probability

softmax_prob <- function(scores, alpha) {

 exp_scores <- exp(alpha * scores)

 return(exp_scores / sum(exp_scores))

}
```

```
# Shannon entropy
calculate_entropy <- function(x, bins = 20) {
 if (length(x) < 2) return(0)
 hist_counts <- hist(x, breaks = bins, plot = FALSE)$counts
 hist_counts <- hist_counts[hist_counts > 0]
 probs <- hist_counts / sum(hist_counts)
 return(-sum(probs * log(probs)))
}

# Corrected simulation function
simulate_east_corrected <- function(N_initial = 10000,
                                    T = 100,
                                    alpha = 2.0,
                                    beta = 0.4,
                                    threshold = 0.3,
                                    seed = 42) {

 set.seed(seed)

 # Initialize included group
 included_strategies <- runif(N_initial, 0, 1)

 # Excluded group starts empty
 excluded_strategies <- numeric(0)

 # Storage
 included_diversity <- numeric(T)
 excluded_diversity <- numeric(T)
```

```
included_mean_utility <- numeric(T)
excluded_mean_utility <- numeric(T)
included_size <- numeric(T)
excluded_size <- numeric(T)
utility_gap <- numeric(T)

for (t in 1:T) {
 # Record current state
 if (length(included_strategies) > 0) {
  included_diversity[t] <- calculate_entropy(included_strategies)
  included_mean_utility[t] <- mean(included_strategies)
  included_size[t] <- length(included_strategies)
 } else {
  included_diversity[t] <- 0
  included_mean_utility[t] <- 0
  included_size[t] <- 0
 }

 if (length(excluded_strategies) > 0) {
  excluded_diversity[t] <- calculate_entropy(excluded_strategies)
  excluded_mean_utility[t] <- mean(excluded_strategies) - beta
  excluded_size[t] <- length(excluded_strategies)
 } else {
  excluded_diversity[t] <- 0
  excluded_mean_utility[t] <- 0
  excluded_size[t] <- 0
 }
```

```
utility_gap[t] <- included_mean_utility[t] - excluded_mean_utility[t]

if (length(included_strategies) == 0) break

# Compute proxy scores
proxy <- included_strategies ^ 2
probs <- softmax_prob(proxy, alpha)

# Selection: probabilistic
selected_mask <- rbinom(length(included_strategies), 1, probs) == 1

# Selected: stay in included (with variation)
new_included <- included_strategies[selected_mask]
variation <- rnorm(length(new_included), 0, 0.05)
new_included <- pmax(0, pmin(1, new_included + variation))

# Rejected: move to excluded
new_excluded <- included_strategies[!selected_mask]

# Update
included_strategies <- new_included
excluded_strategies <- c(excluded_strategies, new_excluded)

# Prevent extinction
if (length(included_strategies) < N_initial * 0.1) {
  n_restore <- N_initial * 0.2 - length(included_strategies)
  if (n_restore > 0) {
    included_strategies <- c(included_strategies, runif(n_restore, 0, 1))
```

```
    }
   }
  }

  return(list(
   included_diversity = included_diversity,
   excluded_diversity = excluded_diversity,
   included_mean_utility = included_mean_utility,
   excluded_mean_utility = excluded_mean_utility,
   utility_gap = utility_gap,
   included_size = included_size,
   excluded_size = excluded_size
  ))
}

# Run for different beta values
run_comparison <- function(beta_values = c(0, 0.2, 0.4)) {
  results <- list()
  for (beta in beta_values) {
   cat(sprintf("Running simulation for β = %.1f...\n", beta))
   results[[as.character(beta)]] <- simulate_east_corrected(
    N_initial = 10000, T = 100, alpha = 2.0, beta = beta, seed = 42
   )
  }
  return(results)
}

# Execute
```

```
results <- run_comparison()


# Print corrected gap analysis
cat("\n", strrep("=", 70), "\n", sep = "")
cat("CORRECTED PROTECTION PARADOX GAP ANALYSIS\n")
cat(strrep("=", 70), "\n\n", sep = "")


for (beta in names(results)) {
  res <- results[[beta]]
  final_gap <- res$utility_gap[length(res$utility_gap)]
  cat(sprintf("β = %s:\n", beta))
  cat(sprintf("  Utility gap at round 100: %.3f\n", final_gap))
}
```

Appendix C: Measurement Protocol

C.1 PGI Estimation (α)

Given observed selection data ${P(s_i), \text{selected}_i}$, α can be estimated via maximum likelihood:

$$\hat{\alpha} = \arg\max_{\alpha} \prod_{i=1}^{N} \left(\frac{\exp(\alpha P(s_i))}{\sum_j \exp(\alpha P(s_j))} \right)^{\mathbf{1}[\text{selected}_i]}$$

C.2 SSI Construction

$$SSI = \frac{1}{3}(\text{Gini}_{\text{wealth}} + \text{Gini}_{\text{education}} + \text{Gini}_{\text{access}})$$

All components standardized to $[0,1]$. Higher SSI indicates fewer outside options.

C.3 BDD Calculation

For each time window of length W:

$$H_t = -\sum_{k=1}^{K} \hat{p}_{t,k} \log \hat{p}_{t,k}$$

where $\hat{p}_{t,k}$ is the proportion of agents in behavioral state k at time t.

C.4 λ and Gap Acceleration Estimation

Using segmented regression with change point detection at algorithm introduction time τ:

$$\log H_t = \beta_0 + \beta_1 t + \beta_2 (t - \tau)_+ + \varepsilon_t$$

Then $\lambda_{pre} = -\beta_1$ and $\lambda_{post} = -(\beta_1 + \beta_2)$.

For gap acceleration:

$$\Delta_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \varepsilon_t$$

Positive γ_2 indicates acceleration ($\frac{d^2\Delta_t}{dt^2} > 0$).