PIONEER
ACADEMIC PUBLISHING LIMITED

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 4 Number 5 October 2025

Algorithmic Reinforcement and the Co-Evolution of User Preferences Through a Mechanistic Analysis of Diversity Loss in Recommender Systems

K. Nowak[1]

[1] Warsaw University of Technology, Poland
Correspondence: K. Nowak, Warsaw University of Technology, Poland.

doi:10.56397/JPEPS.2025.10.01

Abstract

Recommender systems have evolved into adaptive infrastructures that mediate human attention, learning, and decision-making across digital environments. This paper presents a mechanistic analysis of how algorithmic reinforcement processes co-evolve with user preferences, producing a progressive reduction in informational diversity. By conceptualizing recommendation as a coupled dynamical system, the study explains how reinforcement learning architectures internalize behavioral feedback and transform transient user interactions into long-term preference structures. The analysis identifies a recursive mechanism in which both algorithmic policies and user cognition adapt toward equilibrium states that maximize predictability and engagement at the expense of novelty. Empirical findings and theoretical models from recent reinforcement learning research are synthesized to elucidate the dynamics of diversity loss as an emergent property of co-adaptation. The paper proposes a mechanistic framework that integrates stochastic exploration, entropy regularization, and temporal reward shaping to sustain informational variety in reinforcement-driven ecosystems. This approach reconceptualizes recommender systems as co-evolutionary environments where the preservation of diversity is a structural necessity for epistemic resilience, cognitive openness, and sustainable engagement.

Keywords: algorithmic reinforcement, recommender systems, diversity loss, informational entropy, reinforcement learning

---

1. Introduction

Recommender systems shape much of contemporary digital experience by filtering, ranking, and suggesting information according to individual behavioral traces. These systems translate human activity into quantifiable signals that inform future algorithmic decisions, producing a recursive cycle of interaction between user and model. Algorithmic reinforcement emerges when these cycles continuously adjust recommendations based on observed responses, embedding the history of user engagement into the architecture of the system itself. The mechanism transforms transient preferences into persistent behavioral patterns, aligning exposure to content that maximizes predicted satisfaction or attention. Over repeated iterations, the recommender's optimization objectives begin to steer the evolution of user tastes, creating a co-dependent dynamic in which algorithms both reflect and construct preference landscapes.

As this interaction intensifies, a structural narrowing of informational diversity occurs. The user's choice space becomes progressively confined to familiar categories because the system interprets consistency as preference certainty. Engagement-driven optimization accentuates this contraction, since content that diverges from prior patterns yields lower immediate reward signals. The loss of diversity manifests not only in the reduction of novel or unexpected items but also in the attenuation of cognitive variety, where users encounter fewer distinct perspectives or genres. Such homogenization represents a form of algorithmic path dependence, where early behavioral signals disproportionately influence long-term exposure trajectories.

A mechanistic understanding of this process requires examining the interplay between learning algorithms, feedback design, and user cognition. Reinforcement learning agents operate through the maximization of cumulative reward functions, while users respond through evolving attention and interest parameters. Each click, skip, or linger constitutes a data point in a continuous adaptive system. Over time, the optimization of engagement metrics induces emergent properties that are not explicitly programmed, such as polarization or preference ossification. The recommender thus becomes an active participant in shaping the informational environment, operating within a feedback architecture that couples computational efficiency with human behavioral plasticity.

2. Algorithmic Reinforcement Dynamics

At the algorithmic level, reinforcement learning has become central to the design of modern recommender systems because it provides a structured paradigm for modeling sequential decision-making under uncertainty. In this setting, each recommendation constitutes an *action*, and each user response—whether a click, dwell, or skip—serves as a *reward* that updates the system's policy. The recursive adaptation of the algorithm produces a continuous feedback loop in which the model both learns from and reshapes user behavior. This process has been extensively documented in reinforcement learning–based recommender surveys such as Afsar et al. (2022), which describe how engagement-maximizing policies tend to overemphasize short-term gains while diminishing exposure diversity.

The core of algorithmic reinforcement lies in the definition of its reward structure. Conventional designs prioritize immediate engagement signals, often ignoring the long-term diversity of user experience. Such optimization biases can induce *mode collapse*, a state in which the system repeatedly selects from a narrow set of high-reward items. In studies by Lin et al. (2023), actor-critic architectures used in deep reinforcement learning recommender systems exhibit strong tendencies toward reward overfitting, where the model reinforces repetitive exposure to previously popular content. These findings illustrate how algorithmic reinforcement transforms localized behavioral regularities into systemic recommendation biases.

A deeper mechanistic understanding requires examining how different reinforcement learning formulations affect diversity outcomes. Policy-gradient algorithms adjust parameters by estimating gradients of expected returns, while value-based methods such as Deep Q-Networks assign numerical value to discrete recommendation options. Although both methods optimize long-term user satisfaction in theory, in practice they gravitate toward strategies that replicate familiar interaction patterns. Zou et al. (2019) show that optimizing for cumulative engagement without temporal discounting leads to reinforcement of repetitive behaviors rather than exploration of new content spaces. This effect demonstrates that algorithmic reinforcement, when guided by narrowly defined reward signals, tends to amplify preexisting behavioral biases rather than diversify them.

Architectural features within deep models further intensify this convergence. Embedding-based recommendation systems learn latent representations that capture correlations among users, items, and contextual factors. When reinforcement updates operate within these compressed manifolds, the model gradually collapses onto regions of the latent space that correspond to high-probability, high-reward interactions. Studies such as Yu et al. (2024) propose modified reward structures that integrate diversity and novelty metrics directly into the reinforcement objective, demonstrating measurable improvements in content variety without substantial losses in predictive accuracy.

Algorithmic reinforcement dynamics therefore encompass not only a computational optimization process but a co-evolutionary mechanism linking machine learning parameters to human behavioral tendencies. Each iteration of feedback stabilizes mutual expectations between system and user, reducing uncertainty while constraining variability. The emergent equilibrium reflects a convergence of algorithmic learning and human preference formation. Diversity loss thus becomes a structural byproduct of adaptive optimization—a form of informational entropy reduction within the human–algorithm ecosystem, as also noted by Zhao et al. (2025) in their comprehensive analysis of fairness and diversity trade-offs in recommendation algorithms.

In practical implementations, deep reinforcement learning architectures such as actor–critic models and policy-gradient algorithms demonstrate strong sensitivity to the structure of historical data. These architectures optimize policies by evaluating the expected return of each action sequence, yet their dependence on previously observed reward distributions produces a narrowing of exploration space. The agent learns to exploit the most predictable regions of user behavior because those regions offer the highest cumulative reward signal. When applied to recommender systems, this process intensifies the repetition of familiar content and suppresses the discovery of less represented items. Lin et al. (2023) identify this as a manifestation of overfitting within the reinforcement framework, where the model's learned policy becomes confined to historical engagement patterns. Such confinement reduces informational entropy in the recommendation process, creating deterministic cycles of exposure that reinforce prior user trajectories.

The mechanism of overfitting in reinforcement recommenders originates in the structure of the reward gradient. Actor–critic models update both policy and value networks based on observed feedback. When user interactions are concentrated in narrow behavioral clusters, the agent's reward landscape becomes steep around these dominant actions. This leads to a disproportionate update magnitude for frequently rewarded items and minimal parameter adjustment for novel or uncertain content. The result is a rapid convergence toward exploitation. Empirical evidence suggests that such convergence diminishes system robustness by amplifying homogeneity in user experiences and weakening long-term engagement stability.

To mitigate this, recent research explores *multi-objective reinforcement learning* frameworks that balance accuracy with novelty and diversity. Stamenkovic et al. (2022) propose auxiliary loss functions that integrate diversity-aware reward terms into the training process. These formulations optimize for both relevance and entropy, allowing the system to maintain recommendation precision while sustaining exposure to new information. By incorporating diversity constraints into the reinforcement objective, the model learns to navigate a broader state–action space, countering the collapse toward high-reward attractors. This adjustment transforms the algorithm's learning trajectory from a purely exploitative regime toward a balanced adaptive system that values informational variety as part of its optimization landscape.

3. Mechanistic Co-Evolution of User Preferences

User preferences evolve through an intricate process of continuous interaction with algorithmic environments. Each recommendation presented to a user represents not only a predictive output of the system but also a behavioral input that shapes subsequent algorithmic decisions. This recursive interaction establishes a coupled feedback mechanism in which human cognition and computational inference co-regulate one another. The user's attention, emotion, and memory interact with the algorithm's reward and policy structures, producing a cycle of mutual adaptation. Empirical research by Möller et al. (2020) shows that personalized recommendation processes in news and media contexts systematically narrow users' informational exposure. The mechanism is not the direct result of deliberate design but arises from the inherent feedback structure of the system itself, where repetitive exposure subtly shifts user expectations of relevance.

The co-evolution of user preferences can be understood as an emergent property of coupled dynamical systems. In such systems, the algorithm's policy update function and the user's preference formation process operate as interacting equations that continuously influence each other's parameters.

Table 1. A Comparative Summary of Algorithmic and Human Adaptation Mechanisms in the Co-Evolution Process

Mechanistic Dimension	Algorithmic Adaptation (System)	User Adaptation (Human)	Emergent Effect
Learning Rule	Gradient-based policy update	Cognitive reinforcement through exposure	Feedback amplification
Objective Function	Maximization of engagement reward	Minimization of cognitive effort	Mutual predictability
Memory Dynamics	Weighted by recent interactions	Shaped by repetition and familiarity	Preference consolidation
Exploration Tendency	Decreases over optimization cycles	Decreases as novelty perception declines	Reduced informational entropy
Adaptation Speed	High (batch or real-time updates)	Moderate (habit formation)	Temporal coupling and equilibrium

Each round of interaction—comprising recommendation, evaluation, and behavioral response—modifies the internal state of both components. The recommender updates its policy based on perceived engagement value, while the user's perceptual schema adapts to the regularities of exposure. Over successive iterations, the joint system begins to self-organize into stable attractor states characterized by high predictability and low diversity. These attractors represent equilibrium configurations where the mutual reinforcement between algorithm and user produces minimal uncertainty.

Cognitively, this process manifests through phenomena such as confirmation bias and habituation. When an algorithm repeatedly presents content consistent with established preferences, the user's attentional filters become attuned to those patterns. Novel or dissonant information receives reduced cognitive processing, diminishing the likelihood of behavioral signals that could trigger algorithmic exploration. The recommender interprets this decline in engagement as evidence of irrelevance, which leads to further suppression of novel content. Over time, this creates a form of perceptual canalization, where the boundaries of curiosity and interest contract around the dominant themes reinforced by the system. Research on adaptive personalization suggests that such preference consolidation results from reinforcement contingencies at both the neural and computational levels, as attention and prediction error jointly stabilize around familiar stimuli.

At the system level, co-evolution produces measurable consequences for informational diversity and social epistemology. When aggregated across millions of users, individual preference reinforcement can amplify macro-scale polarization. Distinct clusters of users evolve along separate attractor pathways, each governed by different feedback parameters within the same underlying algorithmic architecture. The emergent segmentation of audiences reflects the self-organizing properties of the coupled human–machine ecosystem. This pattern has been observed in large-scale simulations of user–algorithm interactions, where even neutral recommendation policies lead to polarized equilibria when user adaptation dynamics are included in the model. Studies such as Zou et al. (2019) provide evidence that feedback reinforcement on engagement-driven platforms accentuates divergence between groups while reducing diversity within them.

The mechanistic view of co-evolution reveals that diversity loss is not a secondary artifact but an intrinsic outcome of adaptive feedback. The user's evolving cognitive model and the recommender's learning algorithm share the same optimization direction toward predictability and efficiency. Both agents, human and computational, minimize uncertainty by aligning their internal states. The mutual drive for coherence and relevance transforms the open space of potential experiences into a constrained domain of familiar patterns. Sustaining diversity within this framework requires interventions that intentionally introduce stochasticity or novelty into the recommendation process. The challenge lies in designing adaptive systems that respect user engagement while maintaining informational entropy at levels that support exploration and cognitive flexibility.

4. Diversity Loss and Systemic Implications

Diversity loss in recommender systems emerges as a structural byproduct of optimization processes that favor stability, efficiency, and engagement. When the algorithm repeatedly selects high-reward actions associated with known user preferences, the variance of recommendations gradually decreases. This contraction of the recommendation space leads to a reduction in informational entropy, a measurable decline in the variety and novelty of suggested items. The mechanism mirrors biological systems that overexploit successful survival strategies at the expense of genetic variation. In recommender environments, algorithmic exploitation accelerates convergence toward predictable outcomes. The outcome is a narrowing of exposure that not only limits user discovery but also reshapes collective cultural and informational ecosystems.

The loss of diversity results from the dominance of short-term reward signals embedded within algorithmic objectives. Engagement metrics such as click-through rate or watch duration are often treated as proxies for satisfaction, yet they capture only the surface dynamics of user interaction. Zhao et al. (2025) show that systems optimized solely for engagement tend to produce homogeneous recommendation distributions even when trained on diverse datasets. This occurs because the gradient-based optimization process amplifies frequent feedback patterns, suppressing low-frequency but potentially meaningful signals. Once these patterns dominate the policy space, the algorithm enters a cycle of self-confirmation in which diversity loss becomes an emergent equilibrium rather than a correctable anomaly.

Efforts to reintroduce diversity often rely on post-processing strategies or fairness constraints. Antikacioglu and Ravi (2017) demonstrate that network flow optimization can redistribute exposure across items and categories with minimal sacrifice to predictive performance. Such interventions, while effective in controlled evaluations, operate downstream of the reinforcement loop. They address the visible outcomes of diversity loss without altering its underlying causal mechanisms. Since the root of the phenomenon lies in the co-adaptive relationship between user feedback and algorithmic reward, external rebalancing techniques cannot produce long-term systemic change. The optimization pressure that drives convergence remains active, gradually eroding post-hoc diversity adjustments.

Sustainable mitigation of diversity loss requires integrating diversity directly into the algorithmic reward structure.

Table 2. Taxonomy of Diversity Intervention Strategies in Recommender Systems

Approach Type	Mechanism of Action	Representative Studies	Expected Outcome
Post-processing (Re-ranking)	Adjusts item distribution after recommendation	Antikacioglu & Ravi, 2017	Short-term diversity increase, minimal accuracy loss
Reward Engineering	Embeds diversity into reinforcement objective	Yu et al., 2024	Balanced novelty–relevance optimization
Regularization Techniques	Adds entropy or variance penalties to model loss	Zhao et al., 2025	Prevents over-concentration in latent space
Stochastic	Introduces probabilistic	Zou et al., 2019	Sustains long-term

Exploration	decision noise		user curiosity
Multi-agent Interaction	Models diversity as emergent from multi-user dynamics	Simulated studies in social recommender contexts	Systemic resilience and exposure variety

Yu et al. (2024) propose multi-objective reinforcement learning frameworks that simultaneously optimize for relevance, novelty, and fairness. These models redefine success by assigning explicit value to informational variety. By embedding entropy-related terms into the reward function, the agent learns to associate exploration with positive long-term return. Such approaches extend the conceptual scope of recommendation from accuracy maximization to ecological balance within the informational environment. The resulting policies exhibit greater resilience to preference homogeneity and more stable long-term engagement patterns.

At a systemic scale, diversity loss alters the informational topology of digital ecosystems. The contraction of exposure pathways reduces cross-domain connectivity, weakening the circulation of ideas and diminishing opportunities for serendipitous discovery. When aggregated across populations, these effects manifest as macro-level informational silos. The system's optimization dynamics thereby influence social cognition and collective knowledge formation. Empirical analyses of content recommendation in news and entertainment contexts reveal that diminished diversity correlates with higher polarization and reduced epistemic breadth. The algorithmic pursuit of engagement thus produces a feedback loop that reconfigures cultural attention landscapes into fragmented clusters.

The implications extend beyond user satisfaction or content fairness. Diversity loss reshapes the cognitive ecology of digital societies by altering how individuals encounter novelty and form judgments about relevance. A mechanistic understanding of this process highlights the inseparability of algorithmic design and human behavior in shaping informational diversity. To counteract systemic homogenization, recommender systems must evolve toward self-regulating architectures that sustain variability as an intrinsic property of learning rather than an externally imposed constraint. Embedding stochastic exploration, contextual randomization, and dynamic novelty thresholds into reinforcement learning policies can maintain informational diversity while preserving personalization quality. These adjustments mark a transition from reactive correction to adaptive equilibrium, positioning diversity as a foundational element of algorithmic intelligence rather than a peripheral metric of fairness.

5. Toward a Mechanistic Framework of Co-Evolution

A mechanistic framework of co-evolution between users and recommender systems begins with the recognition that both entities operate as adaptive agents interacting through feedback. The algorithm functions as a learning agent whose policy evolves through exposure to user behavior, while the user adapts cognitively and affectively in response to the system's recommendations. Each interaction modifies both agents' internal states, creating a continuous exchange of information and influence. When modeled mathematically, this relationship can be expressed as a set of coupled dynamical equations, where the gradient of algorithmic policy updates is conditioned on user response distributions, and the evolution of user preferences depends on the temporal structure of exposure. These coupled equations describe a nonlinear system capable of exhibiting emergent phenomena such as stability, oscillation, and collapse, depending on the balance between exploration and exploitation.

The process can be conceptualized as an autocatalytic loop. Algorithmic outputs act as catalysts that accelerate specific forms of user learning, and user actions feed back into the system as new data that reinforce algorithmic adaptation. Over successive cycles, both agents co-adjust toward equilibrium configurations in which mutual predictability is maximized. Within this equilibrium, diversity loss corresponds to an entropic reduction, where the system's state space contracts into narrow basins of attraction. Zhao et al. (2025) describe this contraction as the systemic manifestation of fairness-diversity trade-offs, arising not from explicit bias but from the inherent optimization structure of feedback-driven models. The coupled system thus transitions from an exploratory regime, rich in potential states, to an exploitative equilibrium characterized by low variability and high stability.

From a theoretical perspective, this dynamic mirrors adaptive processes in biological and ecological systems. In evolutionary biology, diversity is maintained through mutation and recombination that introduce stochasticity into reproductive processes. A similar mechanism is required in recommender ecosystems to avoid informational monocultures. Controlled stochastic exploration introduces variability into algorithmic decision-making without compromising predictive efficiency. Research in multi-objective reinforcement learning has demonstrated that integrating randomness as a structural parameter improves long-term adaptability by preventing convergence to narrow attractor basins. Adaptive regularization schemes, inspired by mutation in population dynamics, can function as a diversity-preserving mechanism by continuously perturbing the policy landscape. Such perturbations enable the system to escape local optima and sustain the flow of novel experiences for users.

The mechanistic framework also requires modeling the temporal interdependence between preference formation and recommendation generation. Zou et al. (2019) show that long-term user engagement can be optimized through reinforcement models that discount immediate rewards in favor of delayed satisfaction. Introducing temporal discount factors alters the geometry of the co-evolutionary system, elongating the feedback horizon and promoting sustained diversity in exposure. The inclusion of time-sensitive reward shaping transforms the coupled system from a reactive to an anticipatory mode, where the algorithm seeks trajectories that maintain user curiosity across extended interactions.

A mechanistic framework of co-evolution therefore integrates principles of dynamical systems theory, evolutionary computation, and behavioral modeling into a unified analytic structure. It conceptualizes the recommender ecosystem as a self-organizing field of interactions rather than a one-directional flow from data to prediction. Diversity loss becomes interpretable as a phase transition, in which the feedback architecture of user–algorithm interaction passes from a state of high entropy and flexibility to one of low entropy and rigidity. To sustain equilibrium between adaptability and stability, recommender systems must be designed to maintain dynamic diversity as an endogenous property. Such systems would not merely mitigate bias or increase novelty but would continually regenerate informational variation through self-regulated exploration. This orientation represents a shift from static optimization toward co-evolutionary intelligence, where algorithmic and human learning processes evolve in tandem to preserve both engagement and epistemic openness.

6. Conclusion

Algorithmic reinforcement and the co-evolution of user preferences form an adaptive system characterized by interdependence, feedback sensitivity, and emergent order. The recommender algorithm learns from behavioral traces while simultaneously constructing the conditions under which those behaviors occur. This duality transforms recommendation from a predictive problem into an ecological process, where patterns of attention, curiosity, and exposure evolve through continuous mutual influence. Diversity collapse within such a system represents a structural manifestation of its learning dynamics rather than an incidental outcome. When optimization objectives prioritize short-term engagement, the system's adaptive capacity narrows. The loss of informational diversity signifies a reduction in the system's entropy, a contraction of the possible trajectories through which users and algorithms can co-adapt.

A mechanistic analysis of this phenomenon reveals the intrinsic coupling between algorithmic design and cognitive evolution. Reinforcement learning models define reward functions that encode implicit value systems, shaping how users encounter information and interpret relevance. Behavioral feedback loops translate individual actions into population-level signals that guide model updates. Over repeated iterations, this alignment between machine inference and human response leads to a systemic equilibrium that privileges predictability. Diversity loss thus emerges as a property of stability within the coupled human–algorithm environment. It reflects the success of optimization in achieving coherence while failing to preserve the variability required for exploration and discovery.

Integrating diversity into the core architecture of recommender systems requires a redefinition of optimization itself. Diversity cannot remain an auxiliary metric applied after relevance is maximized. It must be treated as an essential dimension of value alongside accuracy, satisfaction, and fairness. Multi-objective reinforcement learning frameworks demonstrate that diversity can coexist with engagement when treated as part of the reward structure. Systems designed with adaptive stochasticity, entropy regularization, and temporally extended objectives can sustain variation without sacrificing efficiency. Such architectures transform recommendation from a static feedback engine into a dynamic learning ecosystem capable of self-regulation and renewal.

The ethical implications of this transformation extend beyond algorithmic performance. Diversity preservation concerns the integrity of collective cognition and the resilience of digital culture. Recommender systems now act as infrastructure for knowledge formation, shaping what societies learn and how they perceive the world. A loss of diversity within these infrastructures narrows the horizon of public imagination and reduces exposure to alternative perspectives. Mechanistic awareness of co-evolution offers a pathway toward corrective design, emphasizing equilibrium over dominance and plurality over optimization singularity.

Recommender systems that internalize diversity as a structural principle will not simply deliver content more equitably; they will cultivate environments in which users remain open to uncertainty and novelty. Such systems embody an epistemic ethic grounded in balance, where engagement does not preclude exploration and personalization does not erase variation. The co-evolution of human and algorithmic intelligence then becomes a process of shared adaptation that sustains cognitive richness, cultural multiplicity, and long-term informational resilience.

References

Afsar, M. M., Crump, T., & Far, B. (2022). *Reinforcement learning based recommender systems: A survey.* ACM Computing Surveys.

Chen, M., Wang, Y., Xu, C., Le, Y., & Sharma, M. (2021). Values of user exploration in recommender systems. In *Proceedings of the 29th ACM International Conference on Multimedia* (ACM MM).

Hazrati, N. (2023). *Long-term impact of recommender systems on the evolution of users' choice behaviour.* Free University of Bozen-Bolzano Research Repository.

Hazrati, N., & Ricci, F. (2022). Recommender systems effect on the evolution of users' choices distribution. *Information Processing & Management, 59*(6), 103037.

Kunaver, M., & Požrl, T. (2017). Diversity in recommender systems — A survey. *Knowledge-Based Systems, 123*, 154–162.

Lin, Y., Liu, Y., Lin, F., Zou, L., Wu, P., & Zeng, W. (2023). A survey on reinforcement learning for recommender systems. *IEEE Transactions on Neural Networks and Learning Systems*.

Ma, X., Li, M., & Liu, X. (2024). Advancements in recommender systems: A comprehensive analysis based on data, algorithms, and evaluation. arXiv preprint arXiv:2407.18937.

Stamenkovic, D., Karatzoglou, A., & Arapakis, I. (2022). Choosing the best of both worlds: Diverse and novel recommendations through multi-objective reinforcement learning. In *Proceedings of the 16th ACM Conference on Recommender Systems* (RecSys).

Yu, J., Lyu, S., & Chen, P. (2024). Deep Reinforcement Learning for Boosting Individual and Aggregate Diversity in Product Recommendation Systems. In *Proceedings of the 18th ACM Conference on Recommender Systems (RecSys 2024)*. Association for Computing Machinery.

Zhao, Y., Wang, Y., Liu, Y., & Cheng, X. (2025). Fairness and diversity in recommender systems: A survey. *Proceedings of the ACM on Recommender Systems*.

Zou, L., Xia, L., Ding, Z., Song, J., Liu, W., & Yin, D. (2019). Reinforcement learning to optimize long-term user engagement in recommender systems. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD).*