Algorithmic Game Theory
Biased Learning & Informational Lock-In: A Bandit Model of Echo Chambers

Nadine Daum (245963)

January 4, 2026

Introduction

Persistent misinformation and informational lock-in are central challenges in contemporary information environments. Despite unprecedented access to data, individuals frequently converge on false or incomplete beliefs, particularly in settings shaped by algorithmic curation and selective exposure. These dynamics have been widely studied in media economics, where market incentives and platform design shape both information supply and consumption (Mullainathan & Shleifer, 2005; Gentzkow & Shapiro, 2006). These patterns raise a fundamental question: when do adaptive information systems fail to support effective learning, even when individuals are willing to update their beliefs in response to evidence? A common explanation attributes such failures to cognitive limitations, motivated reasoning, or heuristic decision-making. While these mechanisms are undoubtedly relevant, they do not exhaust the set of possible explanations. A large behavioral literature documents how biased inference or categorical cognition can distort belief formation (e.g., Rabin, 2002; Fryer & Jackson, 2008). At the same time, empirical evidence suggests that individuals do update beliefs when exposed to new information, yet often remain systematically misinformed. This tension suggests that learning failures may arise not only from how beliefs are updated, but also from how information is sampled. This project studies whether persistent mislearning can emerge even when agents update beliefs rationally, solely due to biased information exposure. To isolate this mechanism, learning is modeled as a single-agent multi-armed bandit problem in which information sources have stationary but unknown accuracies. Belief updating follows standard Bayesian or no-regret learning rules, while exposure is distorted by an exogenous bias that shapes which sources are observed. This abstraction separates exposure bias from cognitive bias and allows us to evaluate the robustness of rational learning under distorted sampling. Bandit algorithms are treated as normative benchmarks rather than descriptive models of human behavior (Russo & Van Roy, 2018). The analysis therefore asks a systems-level question: if even idealized learners can fail under biased exposure, then learning outcomes may be fragile properties of the information environment rather than of individual cognition. The results show that biased exposure can induce persistent convergence to inferior information sources, despite correct belief updating and stationary environments. Taken together, the findings suggest that improving learning

outcomes in adaptive information systems may require attention to the structure of information exposure itself, rather than relying solely on assumptions about individual rationality or information quality.

Simulation Set-Up

The information environment is modeled as a single-agent, two-armed Bernoulli bandit. Each arm represents an information source with a fixed but unknown probability of producing an accurate signal. One source, denoted H, is higher quality than the other source L, such that

$$p_H > p_L.$$

The environment is fully stationary: source accuracies do not change over time.

In each period $t = 1, 2, \ldots$, the agent selects a source $A_t \in \{H, L\}$ and observes a binary signal

$$X_t \sim \text{Bernoulli}(p_{A_t}).$$

The agent forms and updates beliefs about source accuracy based on observed signals, but does not directly observe p_H or p_L.

To capture selective exposure, source selection is distorted by an exogenous exposure bias. Let $\pi_t(L)$ denote the probability that the learning rule selects the lower-quality source L at time t. Actual exposure probabilities are given by

$$P_t(L) = (1 - \delta)\pi_t(L) + \delta,$$

where $\delta \in [0, 1]$ measures the strength of biased exposure. When $\delta = 0$, source selection is determined solely by the learning rule. When $\delta > 0$, the agent systematically oversamples source L, independent of its true accuracy.

This formulation isolates exposure bias from belief updating and provides a reduced-form representation of selective exposure mechanisms emphasized in media economics and learning from filtered feedback (Swaminathan & Joachims, 2015).

We vary two key features of the environment. First, the accuracy gap

$$\Delta = p_H - p_L$$

governs how informative the environment is. Second, the exposure-bias parameter δ determines the degree of sampling distortion. Together, these parameters characterize when biased exposure interferes with effective learning.

Implementation

Learning is implemented using three standard bandit algorithms: ε-greedy, UCB1, and Thompson Sampling. These algorithms are chosen as canonical exploration-exploitation benchmarks and

are implemented according to their textbook specifications. No algorithm-specific tuning beyond standard parameters is introduced. TS maintains independent Beta priors over source accuracies and updates posterior beliefs using the Beta-Bernoulli conjugate update rule. Both sources are initialized with identical Beta$(1,1)$ priors. UCB1 is initialized with one pull of each arm to ensure well-defined confidence bounds. The ε-greedy algorithm explores with a fixed exploration probability and otherwise selects the empirically best-performing source. Across all algorithms, learning proceeds under identical horizons and evaluation windows. Exposure bias is incorporated after the algorithm selects an action. At each time step, the learning rule first determines a source according to its internal policy. With probability δ, this choice is overridden and the agent is forced to sample source L; with probability $1-\delta$, the algorithm's selected source is observed. This ordering ensures that bias affects sampling rather than belief updating, and that observed rewards are always drawn from the true underlying distribution. Belief updating and empirical estimation are performed correctly in all cases. TS updates posterior beliefs based solely on observed signals, while ε-greedy and UCB1 update empirical means using realized outcomes. The reward-generating process is never distorted, and the agent observes the true outcome whenever a source is sampled. All simulations are run using fixed random seeds to ensure reproducibility. Results are aggregated across independent runs for each parameter configuration in order to characterize probabilistic learning outcomes. The full simulation code is implemented from scratch and is attached in the appendix.

Results: Informational Lock-in

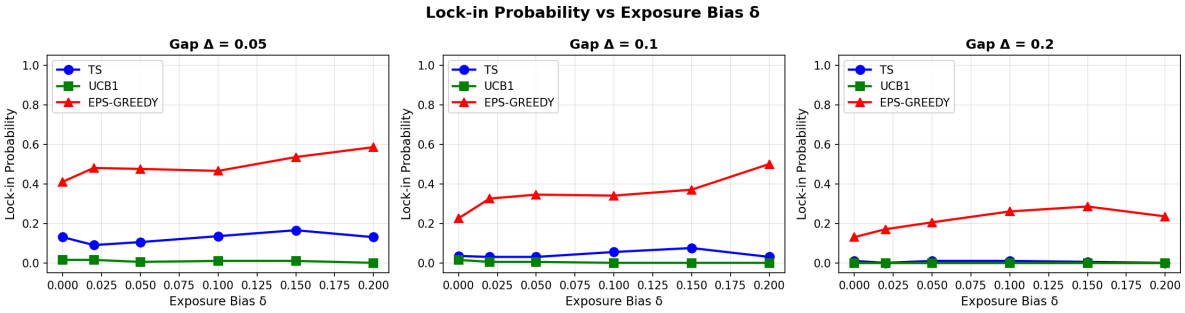

Figure 1: Lock-in probability as a function of exposure bias δ for three bandit algorithms and reward gaps Δ

When exposure is unbiased ($\delta = 0$), all three learning algorithms reliably converge toward the higher-accuracy source H. Thompson Sampling and UCB1 exhibit particularly stable convergence, while ε-greedy displays occasional finite-horizon fluctuations due to persistent random exploration. In the absence of biased exposure, long-run choice shares favor H, cumulative regret remains low, and convergence to the inferior source L is rare. Introducing exposure bias ($\delta > 0$) fundamentally alters learning dynamics. Even though belief updating remains correct and reward distributions are stationary, biased sampling can impede the accumulation of evidence about the superior source. As shown in Figure 1, the probability of informational lock-in increases with δ across all algorithms, with the effect most pronounced when the accuracy gap Δ is small. For sufficiently large exposure bias, agents frequently converge to persistent selection of the lower-quality source L. The severity

of informational lock-in depends jointly on the strength of exposure bias and the informativeness of the environment. Smaller accuracy gaps require only mild bias to induce mislearning, while larger gaps increase robustness and shift the onset of lock-in to higher values of δ. Algorithmic differences are also salient. ε-greedy exhibits high lock-in probabilities even under modest bias, Thompson Sampling substantially mitigates but does not eliminate the risk of lock-in, and UCB1 largely avoids strict lock-in due to forced exploration, albeit at the cost of increased cumulative regret. These patterns reflect strong path dependence in learning under biased exposure. Early biased samples can skew posterior beliefs or empirical estimates, which in turn affect future sampling decisions. Once biased exposure suppresses exploration of the superior source, corrective evidence becomes increasingly unlikely, generating self-reinforcing convergence to the inferior alternative. Informational lock-in thus emerges endogenously from the interaction between rational learning dynamics and distorted information exposure.

Discussion & Policy Implications

The results identify a structural vulnerability of adaptive information systems. Even when users update beliefs rationally and face stationary environments, biased exposure alone can generate persistent mislearning. Learning failures therefore need not stem from cognitive limitations or motivated reasoning, but can arise from the way information is filtered and presented. This mechanism is directly relevant for algorithmic recommendation systems, where content ranking, personalization, and default feeds systematically shape what information users observe. Related work in media economics emphasizes how platform incentives and market forces influence information exposure and belief formation (Mullainathan & Shleifer, 2005; Kasy & Sautmann, 2021). A central trade-off emerges in platform design. Systems that aggressively personalize content may improve short-run engagement or relevance, but risk suppressing informative exploration and thereby increasing the likelihood of informational lock-in. Conversely, mechanisms that promote diversity or exploration can reduce the probability of persistent mislearning, but may entail short-run efficiency or welfare costs. Importantly, these trade-offs arise independently of user irrationality and reflect constraints inherent in adaptive curation. The analysis also clarifies why policies focused solely on improving information quality may be insufficient. Providing access to accurate information does not guarantee effective learning if users are rarely exposed to it. When exposure itself is systematically biased, rational learners may fail to accumulate enough corrective evidence to revise early beliefs, even in the absence of misinformation or deception. From a policy perspective, the results suggest that interventions targeting system-level exposure mechanisms may be more effective than those aimed exclusively at individual cognition. Potential approaches include transparency requirements for recommendation algorithms, constraints on persistent oversampling of aligned but low-quality content, and the use of forced or randomized exposure to introduce controlled exploration. Such interventions can be interpreted as design choices that preserve learning robustness in adaptive information environments. More broadly, the findings highlight a fundamental risk of personalized information systems: personalization can transform otherwise benign learning environments into fragile systems prone to self-reinforcing errors. Rational users alone do not guarantee correct beliefs when the structure of information exposure is endogenously distorted.

Conclusion

From an algorithmic game theory perspective, this project shows how individually optimal learning dynamics can generate collectively undesirable informational outcomes when embedded in biased sampling environments. Even when agents update beliefs correctly and face stationary reward distributions, exposure bias alone can induce path-dependent convergence to inferior information sources. Informational lock-in thus represents a structural failure mode of rational learning systems driven by distorted sampling rather than cognitive limitations. The framework extends to settings with endogenous platforms. In many real-world information environments, exposure bias is not exogenous but generated by strategic recommendation systems that adapt to user behavior. Endogenizing the exposure mechanism would allow analysis of feedback loops between platform objectives, user learning, and long-run belief formation, raising questions about equilibrium selection and welfare in adaptive information markets. The model can also be extended to social learning contexts in which multiple agents learn simultaneously while observing overlapping or correlated information streams. In such settings, biased exposure may propagate across agents, amplifying early sampling distortions and generating collective informational lock-in. Understanding how individual learning dynamics interact with network structure represents a promising direction for future work. Finally, the results motivate a mechanism design perspective on information systems. If biased exposure undermines learning even for rational agents, then platform design can be viewed as the problem of implementing learning-compatible information mechanisms. Designing recommendation systems that balance personalization with exploration, transparency, and robustness to lock-in remains an important and interesting future challenge.

References

Gentzkow, M., & Shapiro, J. M. (2006). Media bias and reputation. *Journal of Political Economy, 114*(2), 280–316. https://doi.org/10.1086/499414

Kasy, M., & Sautmann, A. (2021). Adaptive treatment assignment in experiments for policy choice. *Econometrica, 89*(1), 113–132. https://maxkasy.github.io/home/files/papers/adaptiveexperimentspolicy.pdf

Mullainathan, S., & Shleifer, A. (2005). The market for news. *American Economic Review, 95*(4), 1031–1053. https://doi.org/10.1257/0002828054825619

Russo, D., & Van Roy, B. (2018). A tutorial on Thompson sampling. *Foundations and Trends in Machine Learning, 11*(1), 1–96. https://doi.org/10.1561/2200000070

Swaminathan, A., & Joachims, T. (2015). Counterfactual risk minimization: Learning from logged bandit feedback. In *Proceedings of the 32nd International Conference on Machine Learning* (pp. 814–823).

Appendix: Simulation Code

The full Python implementation used for all simulations is available at `https://github.com/NadineDaum/informational-lock-in-bandits/tree/main`. The repos contains a single Jupyter notebook. The code includes implementations of ε-greedy, UCB1, and Thompson Sampling. All results are fully reproducible. Random seeds are fixed at the run level and varied across independent repetitions. Reported statistics are aggregated across runs to characterize probabilistic learning outcomes.