Boredom-Decay Multi-Armed Bandits as a Sequential Multi-Criteria Decision-Making Framework

Kamil Bortko[0000−0003−3752−3473] and Kacper Fornalczyk[0009−0006−2135−503X] and Jarosław Jankowski[0000−0002−3658−3039]

West Pomeranian University of Technology in Szczecin, Poland
Faculty of Computer Science and Information Technology
http://www.zut.edu.pl
{kbortko,kfornalczyk,jjankowski}@zut.edu.pl

Abstract. In this paper, we introduce Boredom-Decay Multi-Armed Bandits (BD-MAB), a novel extension of the Multi-Armed Bandit (MAB) framework that models boredom as a perceptual–cognitive factor in sequential decision-making. The approach incorporates a boredom-decay mechanism that reduces the perceived value of frequently selected options, capturing habituation and loss of novelty in user-facing systems. From a decision-theoretic perspective, BD-MAB can be viewed as a sequential multi-criteria decision-making (MCDM) framework, combining reward maximization with a dynamic novelty-oriented criterion. Unlike classical MCDM methods with static weights, preference adaptation is embedded implicitly in the decision process. Experimental results show that BD-MAB increases selection diversity (entropy) while maintaining competitive cumulative rewards, supporting the integration of cognitive-inspired mechanisms into bandit algorithms for human-centered systems.

Keywords: multi-armed bandit · multi-criteria decision-making · boredom decay · temporal decision-making · user engagement

1 Introduction

The Multi-Armed Bandit (MAB) problem is a well-known framework for modeling decision-making under uncertainty, balancing exploration and exploitation. Classical MAB algorithms assume stationary rewards determined solely by environmental factors, without considering user perception. In practice, users may experience boredom or perceptual fatigue when repeatedly exposed to the same options [12]. This habituation effect reduces the perceived value of even high-reward options, which is critical in user-facing systems where engagement and satisfaction matter. However, most MAB approaches do not account for such perceptual decay. This motivates incorporating a boredom-aware mechanism into decision-making. In this paper, we use the terms perceptual boredom and boredom-driven perceptual decay interchangeably.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_21

Recent studies [5] show that repeated exposure can lead to user disengagement. While extensions such as rotting bandits model declining rewards [11, 14], they focus on environmental changes rather than internal cognitive effects. Few approaches explicitly incorporate perceptual boredom as an independent factor, highlighting the need for lightweight, boredom-aware strategies.

The main objective of this work is to introduce the Boredom-Decay MAB (BD-MAB), which penalizes frequently selected arms to simulate declining perceived reward. This encourages exploration and prevents premature convergence, improving diversity while maintaining competitive cumulative rewards. We hypothesize that boredom-driven decay increases long-term engagement without significant performance loss.

From a decision-theoretic perspective, real-world systems often involve multiple criteria such as reward, diversity, and user engagement. We interpret BD-MAB within the Multi-Criteria Decision-Making (MCDM) paradigm, where each decision aggregates reward maximization with boredom minimization. Unlike classical MCDM with static weights, BD-MAB adapts criteria dynamically over time, embedding preference evolution implicitly. This positions BD-MAB as a temporal, adaptive MCDM framework bridging reinforcement learning and multi-criteria decision-making.

By modeling boredom as an intrinsic criterion, this work contributes to hybrid MAB–MCDM approaches and demonstrates how cognitive factors can be integrated into sequential decision-making algorithms.

2 Literature Review

The Multi-Armed Bandit (MAB) problem is a fundamental framework for sequential decision-making under uncertainty. Classical algorithms such as ε-greedy, UCB1, and Thompson Sampling balance exploration and exploitation through randomization, confidence bounds, or Bayesian inference [2, 1].

Recent research extends MAB to more complex settings, including adaptive exploration [6], non-stationary environments [13], and contextual or action-centric approaches [7]. Studies have also explored perceptual and behavioral aspects of decision-making [3, 4, 9], highlighting the importance of adapting to user preferences and engagement dynamics [15].

Non-stationary bandit models, such as rotting bandits, explicitly account for reward decay over time [11, 14], encouraging adaptive exploration in dynamic environments. These approaches model changes in environmental rewards but do not capture internal cognitive effects.

In parallel, research in human-computer interaction and cognitive psychology has demonstrated that repeated exposure leads to perceptual habituation and reduced engagement [8]. This phenomenon is critical in user-facing systems, where novelty and diversity influence long-term satisfaction.

Although no prior work explicitly defines the Boredom-Decay MAB (BD-MAB), related approaches address reward decay or large-scale arm spaces [10].

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_21

However, these methods typically focus on external reward dynamics or contextual adaptation rather than internal perceptual effects. Importantly, none model boredom as an independent cognitive penalty applied to arm selection.

To address this gap, we propose BD-MAB, which introduces an intrinsic boredom penalty that increases with repeated selections while preserving the underlying reward distribution. Unlike rotting bandits, the decay is purely perceptual rather than environmental, enabling sustained exploration even in stationary settings.

By integrating perceptual boredom into the MAB framework, this work bridges insights from cognitive science and reinforcement learning, contributing to more human-centered and adaptive decision-making systems.

3 Proposed Method: The Boredom-Decay MAB Algorithm

In this section, we outline the conceptual and mathematical foundations of our proposed algorithm and describe how it differs from classical MAB strategies.

3.1 Classical MAB Strategies

The classical ε-greedy algorithm balances exploration and exploitation by selecting a random arm with probability ε, and otherwise choosing the arm with the highest empirical mean reward. UCB1, on the other hand, uses an optimism-based approach, adding a confidence interval to the estimated rewards to encourage exploration of less frequently chosen arms.

Both of these algorithms treat the reward structure as stationary and do not consider perceptual or behavioral effects that might reduce the subjective value of repeatedly chosen options.

3.2 The Boredom-Decay MAB (BD-MAB) Algorithm

The BD-MAB algorithm extends the ε-greedy framework by introducing a boredom penalty term, $B_k(t)$, which is dynamically updated based on the frequency of choosing arm k. The adjusted expected reward for each arm is calculated as:

$$\tilde{\mu}_k(t) = \hat{\mu}_k(t) - B_k(t),$$

where $\hat{\mu}_k(t)$ is the empirical mean reward for arm k at time t, and $B_k(t)$ represents the boredom penalty, growing with each repeated choice and decaying for unchosen arms:

$$B_k(t+1) = \{\, B_{\,k}\,(t) + \beta, if k = a_t \gamma \cdot B_k(t), otherwise$$

Here, β is the boredom growth rate, and $\gamma \in (0, 1)$ is the boredom dissipation factor.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_21

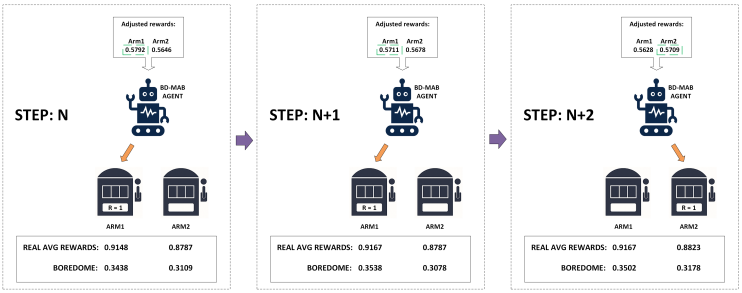

Fig. 1. Conceptual drawing of the algorithm's operation.

This boredom-aware adjustment encourages the algorithm to diversify its selections, reflecting the perceptual reality that repeated exposure to the same option can diminish its perceived attractiveness.

To facilitate understanding, Figure 1 presents a simplified example with two arms over three steps. At each step, the agent selects the arm with the highest adjusted reward, defined as the empirical mean minus the boredom penalty. In round 1, arm 1 is selected due to a higher adjusted reward, yielding reward $R = 1$. The mean reward and boredom vector are then updated. In round 2, arm 1 is chosen again, but its increasing penalty and the decreasing penalty of arm 2 shift the balance. By round 3, arm 2 becomes more favorable and is selected. Importantly, BD-MAB is a lightweight extension that requires only minimal modifications to the classical ε-greedy algorithm and does not rely on external data or user feedback beyond selection history, making it practical and scalable.

3.3 Mapping BD-MAB to Classical MCDM Concepts

From a multi-criteria decision-making (MCDM) perspective, the Boredom-Decay Multi-Armed Bandit (BD-MAB) can be interpreted as a sequential framework where actions are evaluated based on dynamically evolving criteria. While classical MAB focuses on reward maximization, BD-MAB introduces an additional perceptual criterion related to boredom and diversity.

Each arm represents a decision alternative with two implicit criteria: exploitation, modeled by empirical mean reward, and novelty, captured by a boredom penalty that increases with repeated selections and decays over time. Unlike rotting bandits, this penalty reflects a perceptual effect rather than environmental reward changes. BD-MAB operates in a sequential, adaptive setting, where preferences emerge from interaction history. This positions it as a temporal MCDM framework that bridges reinforcement learning with multi-criteria decision analysis in dynamic, human-centered environments.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_21

4 Experimental Setup

To evaluate BD-MAB, experiments were conducted in a stationary MAB environment with 10 arms, each assigned a fixed reward probability sampled from $\mathcal{U}(0,1)$. The focus was on algorithmic behavior and the diversity–reward trade-off.

Algorithm 1: Boredom-Decay Multi-Armed Bandit (BD-MAB)

Input: number of arms n, exploration probability ε, boredom growth rate β, dissipation factor γ
Output: selected arm a_t at each time step
for $i \leftarrow 1$ **to** n **do**
  $\hat{\mu}_i \leftarrow 0$;
  $N_i \leftarrow 0$;
  $B_i \leftarrow 0$;
for *each time step* $t = 1, 2, \ldots$ **do**
  if $rand(0,1) < \varepsilon$ **then**
    Select arm a_t uniformly at random;
  else
    for $i \leftarrow 1$ **to** n **do**
      $\tilde{\mu}_i \leftarrow \hat{\mu}_i - B_i$;
    $a_t \leftarrow \arg\max_i \tilde{\mu}_i$;
  Observe reward r_t;
  $N_{a_t} \leftarrow N_{a_t} + 1$;
  $\hat{\mu}_{a_t} \leftarrow \hat{\mu}_{a_t} + \frac{1}{N_{a_t}}(r_t - \hat{\mu}_{a_t})$;
  $B_{a_t} \leftarrow B_{a_t} + \beta$;
  for $j \leftarrow 1$ **to** n **do**
    if $j \neq a_t$ **then**
      $B_j \leftarrow \gamma \cdot B_j$;

Algorithm 1 extends ε-greedy by introducing a boredom penalty that reduces perceived rewards of frequently selected arms, promoting exploration. The penalty increases for the chosen arm and decays for others, enabling dynamic balance without modifying true rewards.

Experiments used $\varepsilon = 0.1$, $\beta = 0.01$, and $\gamma = 0.99$. Simulations ran for 5000 steps and were repeated 20 times. Performance was evaluated using cumulative reward, moving average reward, and entropy to assess diversity, with ε-greedy as a baseline.

5 Results

Figure 2 shows the cumulative rewards accumulated by BD-MAB and ε-greedy over 5000 time steps. As expected, the ε-greedy algorithm exhibits a slightly

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_21

faster convergence to a higher cumulative reward in these stationary environments. This is due to its direct focus on exploiting the arm with the highest empirical reward without any penalization for repeated choices.

However, achieves competitive cumulative rewards despite its built-in boredom penalty. This indicates that while algorithm encourages exploration, it does not significantly compromise long-term performance in stable scenarios.

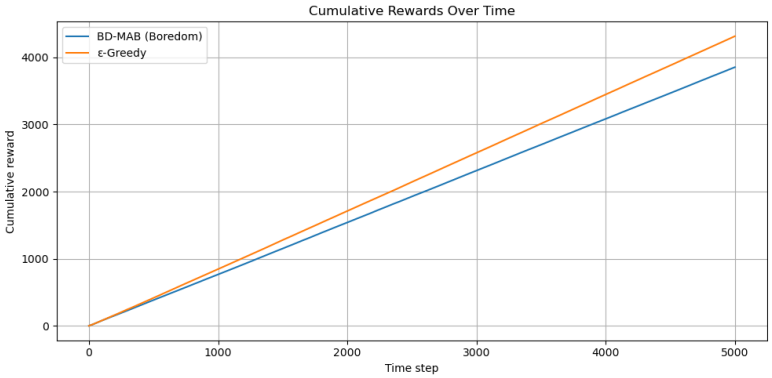

Fig. 2. Cumulative rewards over time for BD-MAB and ε-greedy.

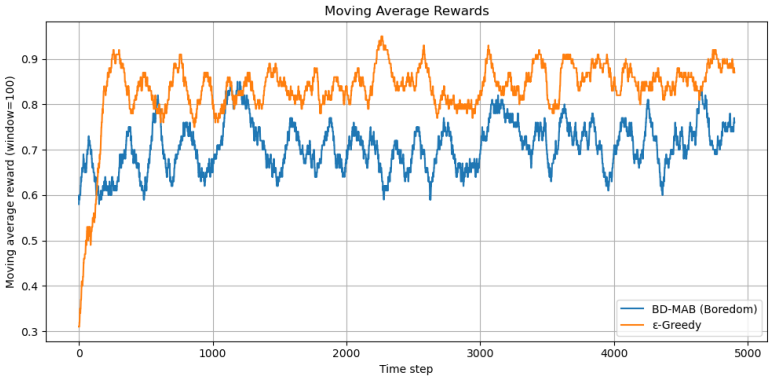

Fig. 3. Moving average of rewards with a window size of 100 time steps. BD-MAB exhibits more fluctuation due to its periodic exploration, stabilizing at competitive performance levels.

Figure 3 presents the moving average of rewards, computed with a window size of 100 time steps. The curves demonstrate that after an initial period of

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_21

increased exploration, BD-MAB stabilizes to a similar moving average as ε-greedy. This suggests that our boredom decay mechanism short-term penalty for repeated choices is gradually balanced out by its adaptive exploration, leading to stable performance in the long run.

To measure the diversity of the arm selections, we compute the entropy of the choice distribution at each time step. Figure 4 reveals that algorithm maintains consistently higher entropy compared to ε-greedy. This confirms that BD-MAB promotes greater diversity in its choices over time.

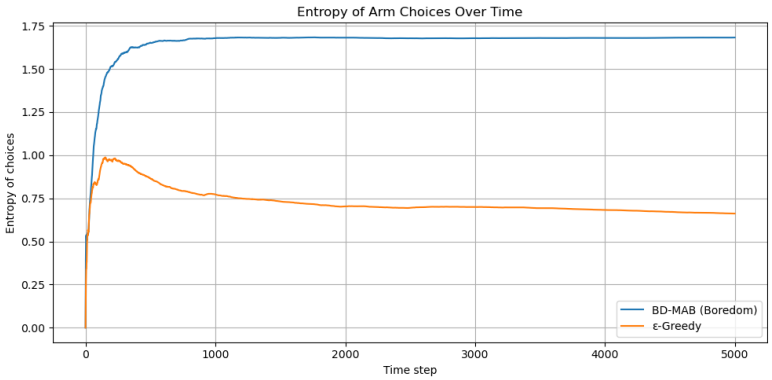

Fig. 4. Entropy of arm choices over time. BD-MAB maintains higher entropy, confirming its ability to sustain selection diversity, which is crucial for user-facing applications.

6 Conclusions

The results demonstrate that BD-MAB consistently achieves higher entropy than the ε-greedy baseline, confirming that the boredom mechanism promotes broader exploration and mitigates premature convergence. This is particularly relevant in user-facing systems, where repeated exposure can reduce engagement. Although BD-MAB attains slightly lower cumulative rewards in stationary environments, this reflects a trade-off between exploration and exploitation, with diversification offering advantages in dynamic settings.

From a decision-theoretic perspective, BD-MAB can be interpreted as a lightweight sequential multi-criteria decision-making framework, combining reward maximization with a dynamic, diversity-oriented boredom criterion. Unlike classical MCDM approaches, criteria are aggregated over time, enabling implicit adaptation without predefined weights.

The method is computationally simple and practical but depends on parameters such as the boredom growth rate β and dissipation factor γ, which influence the balance between exploration and exploitation. While effective, the approach may slow convergence in purely stationary scenarios.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_21

Future work will focus on adaptive parameter tuning, evaluation in non-stationary and user-driven environments, and extensions to contextual and Bayesian bandits, further enhancing the applicability of boredom-aware decision frameworks.

References

1. Agrawal, S., Goyal, N.: Analysis of thompson sampling for the multi-armed bandit problem. Conference on Learning Theory (2012)
2. Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multiarmed bandit problem. Machine Learning **47**(2-3), 235–256 (2002)
3. Bortko, K., Bartków, P., Jankowski, J.: Modeling the impact of habituation and breaks in exploitation process on multi-armed bandits performance. Procedia Computer Science **225**, 4730–4739 (2023)
4. Bortko, K., Fornalczyk, K., Jankowski, J.: Integrating habituation effects with ucb and softmax multi-armed bandit algorithms for optimized digital content delivery. In: International Conference on Computational Science. pp. 153–166. Springer (2025)
5. Camerini, A.L., Morlino, S., Marciano, L.: Boredom and digital media use: A systematic review and meta-analysis. Computers in Human Behavior Reports **11**, 100313 (2023)
6. Dubey, R., Griffiths, T.L., Dayan, P.: The pursuit of happiness: A reinforcement learning perspective on habituation and comparisons. PLoS Computational Biology **18**(8), e1010316 (2022)
7. Greenewald, K., Tewari, A., Murphy, S., Klasnja, P.: Action centered contextual bandits. Advances in Neural Information Processing Systems **30** (2017)
8. Hollebeek, L.D., Maslowska, E., Malthouse, E.C.: The role of recommender systems in fostering consumers' long-term platform engagement. Journal of Service Management **33**(4/5), 721–732 (2022)
9. Killian, J., Lalan, A., Mate, A., Jain, M., Taneja, A., Tambe, M.: Adherence bandits (2023)
10. Kim, Y., Jun, K.S., Nowak, R.: Rotting bandits are no harder than stochastic ones. Proceedings of the 39th International Conference on Machine Learning (ICML) (2022)
11. Levine, N., Crammer, K., Mannor, S.: Rotting bandits. Advances in Neural Information Processing Systems **30**, 3077–3086 (2017)
12. Ma, H., Liu, X., Shen, Z.: User fatigue in online news recommendation. In: Proceedings of the 25th International Conference on World Wide Web. pp. 1363–1372 (2016)
13. Mintz, Y., Aswani, A., Kaminsky, P., Flowers, E., Fukuoka, Y.: Nonstationary bandits with habituation and recovery dynamics. Operations Research **68**(5), 1493–1516 (2020)
14. Seznec, J., Menard, P., Lazaric, A., Valko, M.: A single algorithm for both restless and rested rotting bandits. In: Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics. pp. 3784–3794 (2020)
15. Slivkins, A., et al.: Introduction to multi-armed bandits. Foundations and Trends in Machine Learning **12**(1-2), 1–286 (2019)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_21