Integrating Habituation Effects with UCB and Softmax Multi-Armed Bandit Algorithms for Optimized Digital Content Delivery

Kamil Bortko[0000−0003−3752−3473] and Kacper Fornalczyk and Jarosław Jankowski[0000−0002−3658−3039]

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

Abstract. In the realm of Multi-Armed Bandits (MAB), the integration of habituation with intermittent breaks emerges as a compelling paradigm to enhance exploration-exploitation trade-offs. This study thoroughly investigates the application of habituation with breaks in two prominentstrategies: Softmax and Upper Confidence Bound (UCB). Empirical findings indicate that habituation with breaks can enhance long-term performance, mitigate reward stagnation, and support continuous adaptation. By elucidating the nuanced interplay between habituation, breaks, and MAB strategies, this work aims to inform future developments in decision-making algorithms designed for dynamic, real-world applications.

Keywords: Habituation · Multi-Armed Bandit (MAB) algorithms · Computational

1 Introduction

Multi-Armed Bandit (MAB) algorithms play a pivotal role in decision-making under uncertainty, finding broad applications in fields such as online advertising, clinical trials, and reinforcement learning. The key challenge faced by MAB strategies is balancing exploration (gathering information about untested options) with exploitation (maximizing rewards based on existing knowledge) [4][5]. Recent studies have introduced the promising concept of combining habituation with pauses, aimed at addressing this fundamental dilemma.

Habituation is a well-documented learning phenomenon characterized by a reduction in response to repeated stimuli over time [20]. In the context of MAB algorithms applied to the optimization of visual content delivery, habituation can significantly influence their performance. For instance, if an agent habituates to a suboptimal option (referred to as an arm), it may persistently select this option, ignoring better alternatives [8]. Habituation can be viewed as a form of "forgetting," where the agent gradually loses confidence in its reward estimates for specific options over time.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

Classical MAB research proposes various strategies— epsilon-greedy, Thompson sampling, Softmax, UCB - to balance exploration and exploitation. However, habituation has not been widely studied in this domain (Greenewald et al., 2017). Approaches to avoid stagnation or reward plateaus sometimes include: Entropy-based exploration, Adaptive temperature (for Softmax) and Contextual bandits. While these strategies can mitigate stagnation, none explicitly model habituation as a cognitive-like "forgetting" process. Comparing our approach with such methods could illuminate which strategy is most effective across different non-stationary settings.

In our approach, we propose employing a habituation model that periodically updates estimates while exploiting the option with the highest predicted reward probability. Previous studies [6] have demonstrated the effectiveness of introducing pauses in the context of epsilon-greedy algorithms. We laid the groundwork for understanding how habituation and strategically timed breaks influence the performance of Multi-Armed Bandit (MAB) algorithms [3][2]. That research introduced a novel perspective by integrating behavioral concepts into algorithmic design, revealing the potential benefits of periodic breaks in mitigating habituation effects. Building upon these findings, this article extends the discussion by delving deeper into the interplay between habituation and breaks, with a particular focus on how these mechanisms impact the efficiency of UCB and Softmax strategies. The Softmax strategy introduces a probabilistic element, determining the probability of selecting each arm based on its estimated value. Conversely, the UCB strategy relies on a confidence-based approach, guiding decisions through uncertainty assessments [16].

Integrating habituation with pauses into these strategies may enhance their adaptability and decision-making efficiency, opening new avenues for addressing complex decision-making scenarios [1]. Therefore, our study aims not only to examine how pauses impact these strategies but also to provide practical guidelines for their application in real-world scenarios.

2 Literature review

The existing literature on Multi-Armed Bandit (MAB) algorithms reveals insufficient attention to the topic of habituation. While MAB studies are extensive, the integration of habituation with pauses remains largely unexplored. Habituation, a form of non-associative learning, is defined as the gradual reduction in sensitivity to repetitive, unchanging stimuli [12]. This phenomenon is widespread across various organisms, including humans, and is considered an adaptive mechanism that filters out less significant stimuli to focus resources on novel or important information [18].

In psychological terms, habituation is the process by which an organism reduces its responsiveness to repetitive, non-reinforcing stimuli over time. In this study, we align habituation with the concept of dynamic forgetting: the algorithm gradually loses confidence or "interest" in the same arm if it is repeatedly chosen. As noted by Canadian psychologist Donald Hebb: "Habituation is the

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

first and most elementary form of learning," forming the foundation for more complex learning processes, such as classical or instrumental conditioning [10] [19]. Habituation can be described as a diminishing response to repeated stimuli or a persistent choice of the same arm. Forgetting is an algorithmic mechanism that mimics diminished sensitivity to rewards, which can be reset or partially offset by breaks. By clarifying these notions, we underscore that habituation is not merely a random decay but rather a structured decline in arm preference, potentially mitigated by breaks. Habituation is characterized by a decrease in sensitivity to a given stimulus and the recovery of this sensitivity after a period without exposure [17]. This phenomenon has been extensively studied in fields such as neuroscience, psychology, and biology to better understand learning, memory, attention, and neural plasticity [14] [13].

The Softmax strategy facilitates exploration through probabilistic selection, where the probability of choosing a specific arm is proportional to its estimated reward, taking a parameter into account [7]. A high parameter value results in more exploration, while a low value leads to more exploitation [21]. This probabilistic approach counters habituation by ensuring that even suboptimal arms have a non-zero probability of being selected [22].

UCB algorithms balance exploration and exploitation by selecting arms with the highest upper confidence bound of estimated rewards [4]. This approach ensures more frequent selection of arms with higher uncertainty (less explored), encouraging exploration [9]. The UCB algorithm adjusts the confidence bound based on the number of times an arm has been chosen, reducing the likelihood of suboptimal exploitation over time [1][11].

Observations suggest that the issue of habituation in the context of MAB remains insufficiently investigated, despite its potential benefits for solving decision-making problems under uncertainty. Existing studies on Softmax and UCB strategies also inadequately address this issue, leaving a gap in understanding how habituation with pauses impacts the effectiveness of these strategies. Introducing habituation management strategies is crucial for the effective functioning of MAB algorithms [15]. Both Softmax and UCB offer robust frameworks for balancing exploration and exploitation, reducing the risk of habituation [10]. Future research should develop these methods further and explore their applications across various domains to deepen the understanding of habituation principles in adaptive learning systems.

3 Conceptual framework

Upper Confidence Bound (UCB) and Softmax strategies represent two distinct approaches to the Multi-Armed Bandit (MAB) problem, where an algorithm must effectively select from available options (arms) to maximize rewards.

The application of a habituation model with pauses in UCB and Softmax strategies is justified for several reasons. Firstly, habituation introduces flexibility into the learning process, which is crucial in dynamic environments. Pauses

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

in habituation allow for periodic rest, preventing premature convergence to suboptimal solutions or excessive exploration.

Habituation in MAB algorithms can be conceptually represented as a process in which the agent becomes less sensitive to rewards from a specific arm when it is repeatedly selected. This effect is illustrated in Figure 1 by a continuous curve with a red segment labeled as A. The normal behavior of the algorithm under ideal conditions is reflected in Figure 1 by line B. The occurrence of a break results in an increase in the agent's responsiveness (green curve). The behavior and choices of the agent are shaped by the rewards received, and as the rewards from a specific arm become more predictable, the agent tends to explore other arms that may offer higher rewards.

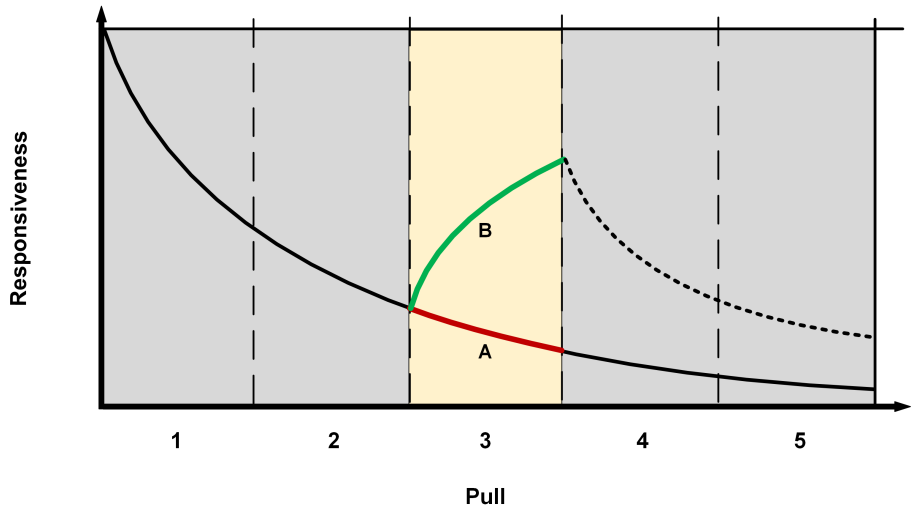

Fig. 1. Example depicting the shape of the responsiveness decline curve with continuous exposure (**A**) and with intermittent exposure breaks (**B**).

It is worth noting that habituation with pauses may be particularly beneficial in situations where the decision-making environment changes or contains unpredictable elements. Habituation allows the algorithm to adapt to changing conditions, minimizing the risk of becoming stuck in suboptimal solutions. Consequently, introducing a habituation model with pauses into UCB and Softmax strategies may enhance their ability to effectively solve dynamic MAB problems.

Furthermore, integrating the habituation model with pauses into UCB and Softmax strategies adds a valuable element of adaptability. Habituation, as a learning phenomenon associated with decreasing sensitivity over time, enables the algorithm to dynamically adjust its responses to changing circumstances. This adaptability becomes crucial in the face of uncertainty or shifts in the reward structure associated with each arm. In the UCB strategy, where the al-

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

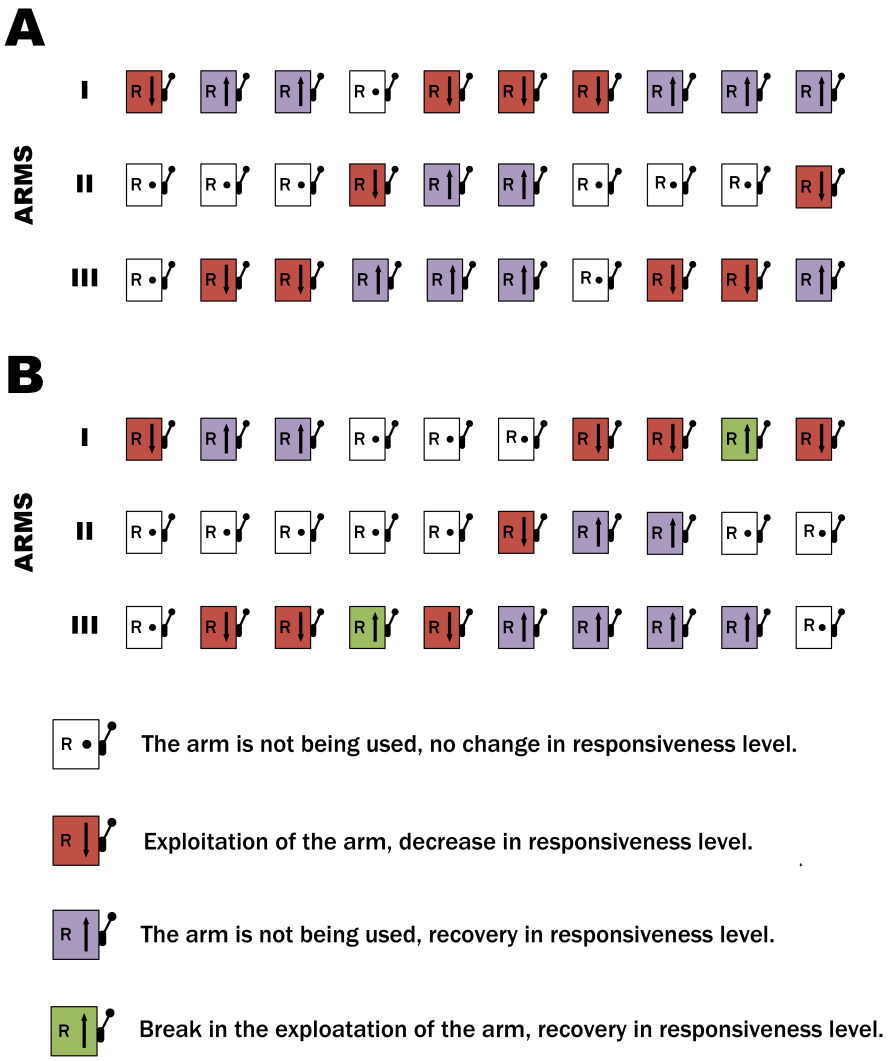

Fig. 2. A graphic illustrating the performance of a MAB algorithm in our experiment with the added effect of habituation,with continuous exposure (**A**) and with intermittent exposure breaks (**B**).

gorithm balances exploitation of known high-reward options with exploration of potentially better ones, habituation with pauses provides a mechanism to prevent premature convergence. Pauses allow the algorithm to periodically reassess its choices, avoiding excessive focus on specific arms and potentially missing better alternatives.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

In the case of the Softmax strategy, which introduces a probabilistic element to decision-making, habituation with pauses supports more balanced exploration of arms. Periodic pauses offer moments of reflection, allowing the algorithm to rethink its exploration-exploitation strategy and adjust probabilities accordingly. This adaptability enhances the Softmax strategy's ability to navigate complex decision spaces and discover optimal solutions.

In summary, applying a habituation model with pauses in UCB and Softmax strategies not only introduces adaptability and flexibility but also addresses specific challenges related to premature convergence and exploration-exploitation balance. This integration increases the algorithms' potential to perform effectively in dynamic decision-making environments, making them valuable tools in areas such as online recommendation systems or autonomous agents facing complex decision scenarios.

4 Experiment Setup

Figure 3 illustrates a segment of our experimental findings, showcasing the average number of selections favoring the highest-reward option over iterations, incremented by 1000, and spanning a total of 3000 iterations. These results provide insight into how the algorithm adapts to different conditions over time.

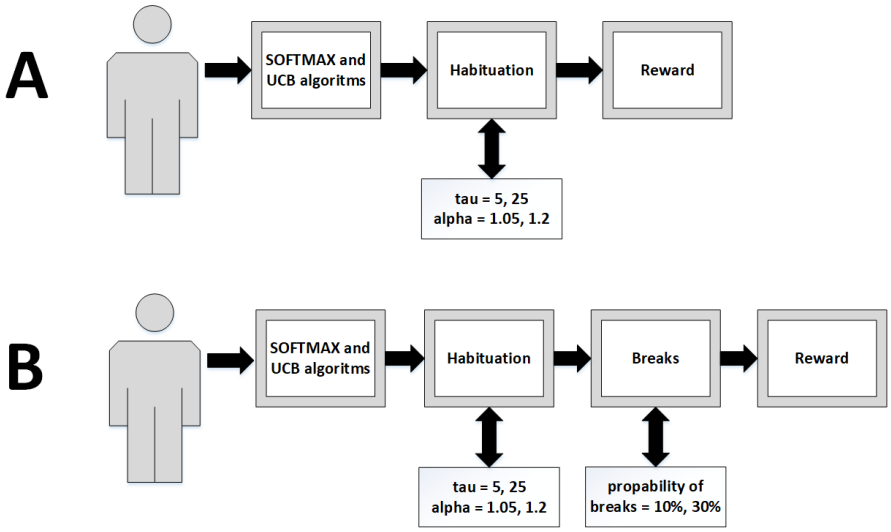

Fig. 3. Experimental approaches: (**A**) performance under habituation with continuous exposure, and (**B**) performance with habituation incorporating breaks.

In Figure 2, we compare the performance of a Multi-Armed Bandit (MAB) algorithm under two experimental conditions: continuous exposure (**A**) and intermittent exposure with breaks (**B**). Figure 2 (**A**) highlights the behavior of the

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

algorithm under continuous exposure, which incorporates the effects of habituation. The analysis reveals that continuous exposure leads to a progressive decline in sensitivity to rewards from specific arms, resulting in suboptimal exploration and exploitation behaviors. This diminishing responsiveness to predictable rewards adversely affects the algorithm's ability to maximize the total reward, highlighting the challenges posed by habituation in decision-making processes. Understanding these dynamics is essential for developing strategies to counteract habituation and improve the performance of MAB algorithms.

Conversely, Figure 2 (**B**) illustrates the performance of the same algorithm when habituation effects are combined with intermittent exposure breaks. The introduction of such breaks markedly enhances the total reward obtained compared to the uninterrupted approach. Periodic breaks mitigate the decline in sensitivity caused by continuous exposure, allowing for more effective exploration and exploitation of available options. This adjustment not only optimizes the algorithm's performance but also demonstrates the potential benefits of integrating controlled interruptions in practical applications. By resetting sensitivity levels, these breaks improve adaptability and mitigate the adverse effects of habituation, ultimately enhancing the efficiency of decision-making in MAB scenarios.

High-Level Procedure illustrating how habituation and "breaks" are introduced into a standard Multi-Armed Bandit (MAB) feedback loop:

1. **Initialization:** Set reward estimates $Q_i(0)$ for each arm i. Initialize any required habituation parameters.
2. **Arm Selection:** Choose an arm i based on either Softmax probabilities or UCB indices.
3. **Reward & Update:** Observe the reward R_i; update Q_i accordingly.
4. **Habituation Decay:** Gradually reduce the sensitivity to repeated pulls of the same arm.
5. **Check Break Condition:** If a break is triggered (e.g., after $X\%$ of trials or upon meeting a certain performance threshold):
   - Partially reset the habituation level.
   - Optionally adjust Q_i or the counters N_i.
6. Repeat until the experiment ends.

The experimental setup for evaluating the Upper Confidence Bound (UCB) algorithm under these conditions involves a systematic comparison between two variants: UCB without breaks and UCB with habituation breaks. Key parameters varied during the experiments include the exploration parameter (c), habituation coefficient (α), temperature parameter (τ), and the percentage of breaks. Within the MAB environment, n arms are initialized with reward probabilities randomized between 0 and 1, and experiments are conducted across a fixed number of iterations. At each iteration, the arm with the highest UCB index—calculated based on estimated rewards and the number of selections—is chosen, the observed reward is recorded, and the algorithm updates its parameters accordingly. This setup allows us to analyze the impact of habituation breaks on the algorithm's performance while systematically varying key parameters.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

Similarly, the Softmax algorithm is evaluated under analogous conditions, comparing its performance without and with habituation breaks. Here, the parameters subject to variation include the habituation coefficient (α), temperature parameter (τ), and the percentage of breaks. Arms are selected according to the Softmax probability distribution, with probabilities assigned based on the estimated rewards. Observed rewards are used to update these estimates and recalibrate the probability distribution, ensuring that the algorithm adapts to the changing environment over successive iterations.

The findings from these experiments underscore the significant role that habituation breaks play in enhancing the performance of MAB algorithms. By addressing the limitations imposed by continuous exposure, such as diminished sensitivity and suboptimal decision-making, the introduction of controlled breaks fosters more effective exploration-exploitation dynamics. These results not only provide a deeper understanding of habituation effects but also open new avenues for optimizing MAB algorithms in diverse applications, ranging from recommendation systems to adaptive learning environments.

4.1 Upper Confidence Bound (UCB) algorithm with habituation and breaks

For each action i selected based on:

1. Select action i according to the rule:

$$i_t = \arg\max_i \left(Q_i(t) + c \cdot \sqrt{\frac{\ln(t)}{N_i(t)}} \right) \tag{1}$$

Where:

- $Q_i(t)$ - estimated average reward for action i at time t,
- $N_i(t)$ - number of selections of action i until time t,
- c - parameter controlling the balance between exploration and exploitation.

2. After performing action i and receiving reward $R_i(t)$, update the estimated average reward for action i:

$$Q_i(t+1) = \frac{(1-\alpha) \cdot Q_i(t) + \alpha \cdot R_i(t)}{N_i(t)+1} \tag{2}$$

Where:

- α - learning rate parameter.

4.2 Softmax algorithm with habituation and breaks

For each action i selected based on the softmax distribution:

1. Select action i according to the softmax distribution:

$$P(i) = \frac{e^{Q_i(t)/\tau}}{\sum_j e^{Q_j(t)/\tau}} \tag{3}$$

Where:

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

– $Q_i(t)$ - estimated average reward for action i at time t,
– τ - parameter controlling the degree of exploration.

2. After performing action i and receiving reward $R_i(t)$, update the estimated average reward for action i:

$$Q_i(t+1) = \frac{(1-\alpha) \cdot Q_i(t) + \alpha \cdot R_i(t)}{N_i(t)+1} \tag{4}$$

Where:

– α - learning rate parameter.

5 Results

The figures 4 and 5 present the cumulative rewards obtained using the Softmax and UCB algorithms, respectively, under varying conditions of habituation and exposure breaks. Each configuration demonstrates the impact of habituation parameters (α and τ) and the introduction of breaks (10% and 30%) on the algorithms' performance over 3000 iterations.

In both figures, non-habituation serves as a baseline, consistently yielding the highest cumulative rewards. The curves for habituation without breaks show a decline in cumulative rewards due to the progressive decrease in sensitivity to repeated rewards, which negatively affects exploration and exploitation efficiency. This decline highlights the inherent challenges of habituation in decision-making processes.

The introduction of breaks (10% and 30%) significantly mitigates the adverse effects of habituation. For the Softmax algorithm in Figure 4, the 30% break consistently outperforms both the non-break and 10% break setups across all parameter combinations, with noticeable improvements in cumulative rewards. The effects are particularly pronounced in configurations where $\alpha = 1.05$ and $\tau = 25$, followed closely by $\alpha = 1.2$ and $\tau = 5$. This indicates that breaks not only reset sensitivity to rewards but also enhance the algorithm's adaptability to dynamic conditions.

Similarly, Figure 5 reveals that the UCB algorithm benefits significantly from the introduction of breaks. While the non-habituation curve remains superior in absolute cumulative rewards, the 30% break approaches this benchmark closely, especially in scenarios where $\alpha = 1.05$ and $\tau = 25$. The improvement is more substantial in the UCB algorithm compared to Softmax, underscoring the effectiveness of breaks in addressing habituation-related performance degradation.

Comparatively, the 10% breaks show moderate improvements but fail to reach the levels achieved with 30% breaks. These results suggest that longer or more frequent breaks are crucial for optimizing algorithmic performance, especially in environments prone to habituation effects.

In summary, the experiments demonstrate that introducing 30% breaks significantly enhances the cumulative rewards for both Softmax and UCB algo-

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

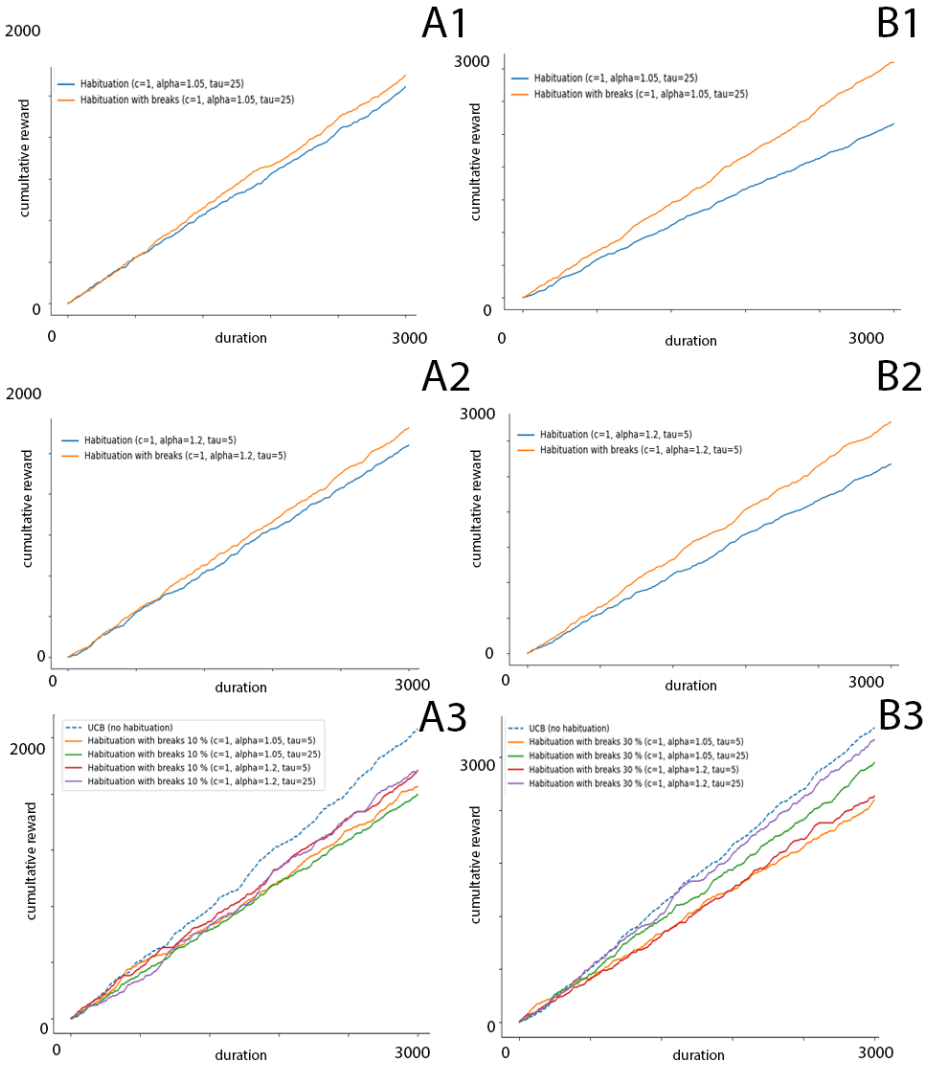

Fig. 4. Performance of the Softmax algorithm under different experimental setups: non-habituation, habituation, and habituation with 10% and 30% breaks.

rithms. The optimal parameter combinations ($\alpha = 1.05, \tau = 25$) further amplify these benefits, providing a robust strategy for mitigating the negative impacts of habituation in Multi-Armed Bandit algorithms. These findings highlight the practical importance of incorporating controlled breaks into algorithmic designs, particularly in applications requiring sustained performance over prolonged decision-making tasks.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

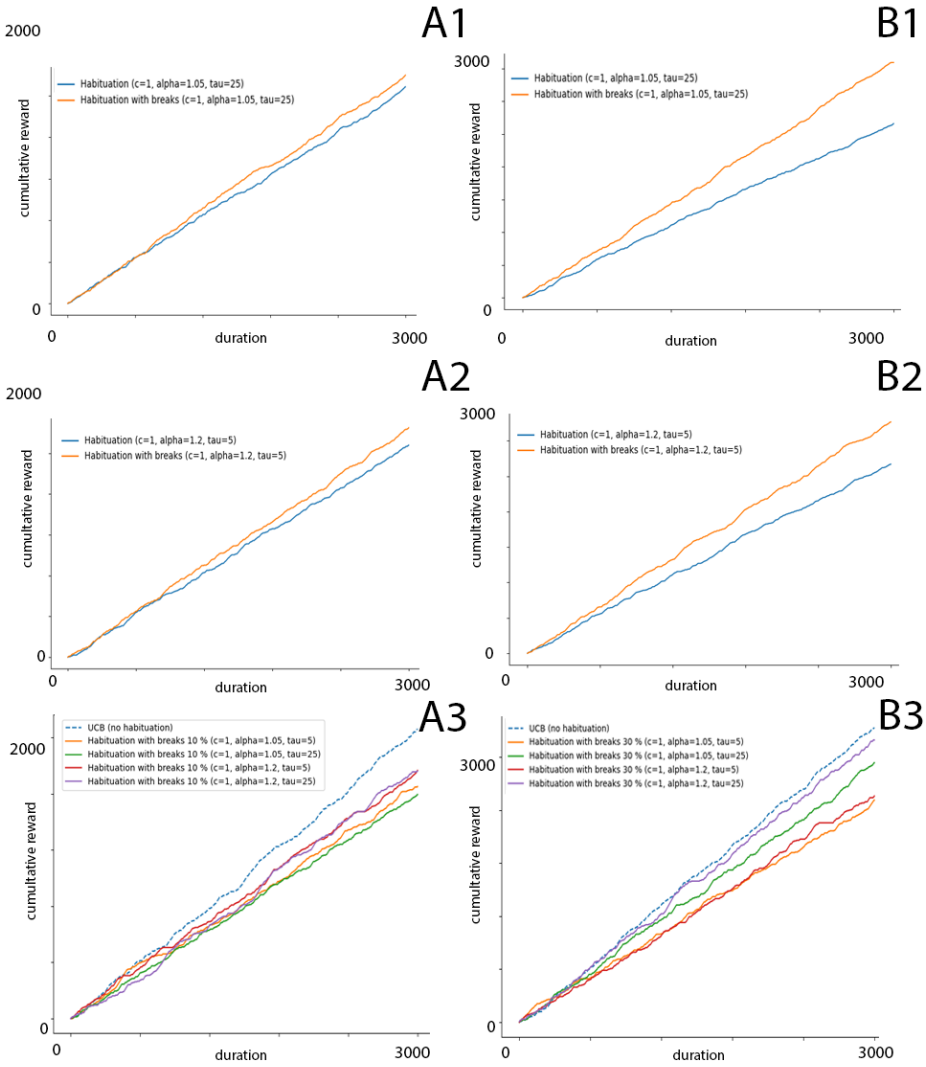

Fig. 5. Performance of the UCB algorithm under different experimental setups: non-habituation, habituation, and habituation with 10% and 30% breaks.

6 Conclusion

Habituation is a multifaceted phenomenon that significantly influences the performance of Multi-Armed Bandit (MAB) algorithms. While habituation poses challenges by diminishing sensitivity to repeated stimuli, incorporating mechanisms such as breaks in exposure and adaptive strategies offers promising solutions. Modeling habituation as a form of "forgetting" enables MAB algorithms

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

to maintain dynamic reward probability estimates, ensuring a balance between exploration and exploitation. This consideration is critical for designing algorithms that achieve optimal outcomes in dynamic environments. The inclusion of breaks in habituation proves particularly advantageous. Firstly, it enhances exploration by encouraging the discovery of potentially more rewarding options, preventing the overexploitation of suboptimal arms. Secondly, breaks mitigate the rapid depletion of high-performing arms, enabling the algorithm to avoid premature convergence and maintain flexibility in decision-making. These benefits are evident in the comparative analysis of Softmax and Upper Confidence Bound (UCB) algorithms. In the Softmax approach, breaks prevent the dominance of a single arm, promoting balanced exploration and exploitation. For UCB, breaks facilitate broader exploration by resetting confidence bounds, leading to improved long-term performance and minimizing the risk of suboptimal convergence. Practical applications, such as recommendation systems, further underscore the importance of habituation breaks. By increasing the diversity of suggested content and preventing repetitive recommendations, breaks can enhance user experience, extend session durations, and boost engagement. Similarly, in advertising platforms, breaks help balance the trade-off between exploring ad effectiveness and exploiting high-performing ads, ultimately maximizing revenue. Beyond these specific domains, the introduction of breaks enhances algorithmic stability and reduces the impact of randomness on results, improving decision precision across various MAB scenarios. Our findings indicate that incorporating breaks into habituation can substantially bolster both exploration and long-term performance in MAB algorithms. By periodically "resetting" diminished responsiveness, the algorithm remains attentive to arms that may have been prematurely overlooked, which is vital in non-stationary environments (e.g., shifting user preferences or volatile market conditions). In the context of potential applications, one can envision a simplified advertising scenario where frequent exposure to the same banner leads to user "ad fatigue" and a drop in click-through rates. By incorporating breaks in the habituation mechanism, the system periodically "refreshes" its selection of ad creatives, thereby avoiding over-promotion of any single campaign. Consequently, this approach maintains higher user engagement, leading to better overall financial performance and a more balanced distribution of ad impressions. This dynamic reactivity translates into more robust, real-world outcomes, such as increased diversity in recommendation systems and improved revenue in advertising platforms. Furthermore, controlled breaks help balance exploitation of currently profitable arms with ongoing exploration, preventing excessive fixation on suboptimal options. Consequently, break-based habituation strategies offer a practical and psychologically inspired means of maintaining adaptability and maximizing cumulative rewards in complex decision-making tasks. Future research and refinement of habituation models in MAB algorithms hold significant potential for advancing both theoretical understanding and practical implementations. By addressing habituation through carefully designed breaks, MAB algorithms can achieve superior

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

outcomes, offering robust solutions for real-world challenges in domains such as finance, advertising, and content recommendation.

7 Future work

Future research on implementing habituation with breaks in Multi-Armed Bandit (MAB) algorithms should focus on several key areas to further enhance the performance of these algorithms and their applications in various fields. One crucial area for future study is the exploration of different parameter combinations, such as tau in Softmax algorithms and alpha in Upper Confidence Bound (UCB) algorithms. Different values of these parameters can significantly affect the performance of the algorithms, so extensive experimentation is necessary to identify the optimal settings. Additionally, while simulations and theoretical analyses provide valuable insights, implementing and testing these algorithms with breaks in real-world scenarios, such as recommendation systems, advertising platforms, and financial applications, is essential for assessing their practical utility. Future work should focus on real-world deployments to understand how these algorithms perform with real data and dynamically changing conditions. Furthermore, exploring the integration of additional contextual information or advanced function approximation techniques, such as neural networks, could lead to more adaptive and efficient exploration strategies. Another promising direction is the development of hybrid approaches that combine break-based habituation with other exploration-enhancement mechanisms, ultimately improving long-term performance across diverse and rapidly evolving application domains.

References

1. Audibert, J.Y., Munos, R., Szepesvári, C.: Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. Theoretical Computer Science **410**(19), 1876–1902 (2009)
2. Auer, P.: Using confidence bounds for exploitation-exploration trade-offs. Journal of Machine Learning Research **3**(Nov), 397–422 (2002)
3. Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multiarmed bandit problem. Machine Learning **47**, 235–256 (2002)
4. Auer, P., Cesa-Bianchi, N., Freund, Y., Schapire, R.E.: The nonstochastic multi-armed bandit problem. SIAM Journal on Computing **32**(1), 48–77 (2002)
5. Bastani, H., Bayati, M., Khosravi, K.: Mostly exploration-free algorithms for contextual bandits. Management Science **67**(3), 1329–1349 (2021)
6. 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)
7. Galichet, N., Sebag, M., Teytaud, O.: Exploration vs exploitation vs safety: Risk-aware multi-armed bandits. In: Asian Conference on Machine Learning. pp. 245–260. PMLR (2013)
8. Greenewald, K., Tewari, A., Murphy, S., Klasnja, P.: Action centered contextual bandits. Advances in Neural Information Processing Systems **30** (2017)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13

9. Grover, A., Markov, T., Attia, P., et al.: Best arm identification in multi-armed bandits with delayed feedback. In: International Conference on Artificial Intelligence and Statistics. pp. 833–842. PMLR (2018)
10. Hebb, D.O.: What psychology is about. American Psychologist **29**(2), 71 (1974)
11. Hillel, E., Karnin, Z.S., Koren, T., Lempel, R., Somekh, O.: Distributed exploration in multi-armed bandits. Advances in Neural Information Processing Systems **26** (2013)
12. Jankowski, J.: Habituation effect in social networks as a potential factor silently crushing influence maximisation efforts. Scientific Reports **11**(1), 19055 (2021)
13. Karnin, Z., Koren, T., Somekh, O.: Almost optimal exploration in multi-armed bandits. In: International Conference on Machine Learning. pp. 1238–1246. PMLR (2013)
14. Lu, T., Pál, D., Pál, M.: Contextual multi-armed bandits. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. pp. 485–492. JMLR Workshop and Conference Proceedings (2010)
15. Mintz, Y., Aswani, A., Kaminsky, P., Flowers, E., Fukuoka, Y.: Nonstationary bandits with habituation and recovery dynamics. Operations Research **68**(5), 1493–1516 (2020)
16. Pike-Burke, C., Grunewalder, S.: Recovering bandits. Advances in Neural Information Processing Systems **32** (2019)
17. Rankin, C.H., Abrams, T., Barry, R.J., et al.: Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. Neurobiology of Learning and Memory **92**(2), 135–138 (2009)
18. Slivkins, A., et al.: Introduction to multi-armed bandits. Foundations and Trends in Machine Learning **12**(1-2), 1–286 (2019)
19. Stanley, J.C.: Computer simulation of a model of habituation. Nature **261**(5556), 146–148 (1976)
20. Thompson, R.F.: Habituation: a history. Neurobiology of Learning and Memory **92**(2), 127 (2009)
21. Thompson, W.R.: On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika **25**(3-4), 285–294 (1933)
22. Zhou, L.: A survey on contextual multi-armed bandits. arXiv preprint arXiv:1508.03326 (2015)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_13