Research on Time Discounting and Magnitude Effect in Intertemporal Decision-Making from the Perspective of Dual-System Theory

Kun Zhang

Ningxia University School ofTeacher Education, Yinchuan, 750021, Ningxia, China

Keywords: intertemporal decision-making, time discounting, dual-system theory, reinforcement learning

Abstract: Intertemporal decision-making, which involves choosing between immediate and delayed outcomes, is shaped by time discounting and the magnitude effect. Traditional models, such as exponential and hyperbolic discounting, often fail to capture the dynamic inconsistencies and reward size effects seen in human behavior. To address these gaps, this paper introduces the Dual-System Reinforcement Learning (DSRL) algorithm, which integrates dual-system theory with reinforcement learning. The DSRL model dynamically adjusts the influence of impulsive and rational decision-making systems based on reward magnitude and time delay. By incorporating a hybrid discounting mechanism, the DSRL algorithm better models how individuals weigh short-term versus long-term rewards. Experimental results demonstrate that DSRL surpasses baseline models in handling time discounting and adapting to different reward magnitudes, achieving higher cumulative rewards over time. Additionally, the DSRL model learns more efficiently, converging faster to optimal decision strategies. These findings suggest that DSRL provides a more accurate and adaptable framework for understanding human decision-making in complex, delayed-reward environments.

1. Introduction

Intertemporal decision-making, the process by which individuals make choices involving trade-offs between immediate and delayed outcomes, is central to fields such as economics, psychology, and behavioral science. A critical concept in understanding these decisions is time discounting, where individuals devalue future rewards relative to more immediate ones [1-2]. For example, a smaller immediate reward is often preferred over a larger future reward, highlighting how time influences decision preferences. Another well-established observation is the magnitude effect, where larger rewards are discounted less steeply than smaller ones [3-4]. Both of these phenomena pose fundamental challenges in modeling how individuals evaluate future outcomes and their trade-offs. A range of models has been developed to explain time discounting behavior. One of the earliest models, exponential discounting, introduced by Samuelson [5], assumes a constant discount rate over time. However, this model often fails to explain real-world behavior, where people exhibit preference reversals, preferring smaller-sooner rewards in the short term but favoring larger-delayed rewards in the long term. To address these inconsistencies, the hyperbolic discounting model was proposed, which assumes that the discount rate decreases with increasing time delays [6-7]. This model better captures dynamic inconsistency, but struggles to explain the influence of reward size, particularly the magnitude effect.

The magnitude effect, where individuals discount larger rewards less than smaller rewards, suggests that the size of the reward affects temporal preferences [8]. While empirical evidence supports this, traditional models like hyperbolic discounting do not fully capture this effect, leading to the need for new approaches that can account for both time discounting and reward magnitude in a cohesive framework. Another key development in understanding intertemporal decisions is the dual-system theory, which explains decision-making as the interplay of two systems: System 1, which is fast, impulsive, and emotion-driven, and System 2, which is slower, deliberate, and rational [9]. System 1 tends to favor immediate gratification, while System 2 focuses on long-term consequences. This theory has provided valuable insights into why people exhibit time-inconsistent

  DOI: 10.25236/ieesasm.2024.053

preferences, such as opting for immediate rewards in the short term but favoring delayed rewards in the long term. However, most computational models have yet to fully integrate the dual-system framework with the magnitude effect, leaving a gap in modeling how both temporal and magnitude factors interact in decision-making.

To address these limitations, we propose a Dual-System Reinforcement Learning (DSRL) model that integrates dual-system theory with reinforcement learning to capture both the time discounting and magnitude effects. The DSRL model dynamically adjusts the influence of System 1 and System 2 based on the size of the reward and the time horizon. Unlike traditional models that use fixed discount rates, the DSRL model continuously learns and adapts decision strategies through feedback, providing a flexible and realistic representation of how individuals balance short-term impulsivity with long-term rationality. The DSRL model offers several contributions. First, it unifies the understanding of time discounting and the magnitude effect within the context of dual-system theory. Second, the model's dynamic adjustment of decision-making processes allows for more accurate predictions across different decision contexts. Finally, its reinforcement learning mechanism enhances adaptability and robustness, making it more reflective of real-world decision-making. This novel approach helps fill existing gaps in the literature and offers new insights into the cognitive processes that govern intertemporal choices.

2. Methodology

2.1 Dual-system reinforcement learning

The Dual-System Reinforcement Learning (DSRL) algorithm is designed to integrate the dual-system theory of decision-making with reinforcement learning techniques. In this framework, human decision-making is influenced by two competing systems: System 1, which is impulsive and driven by emotions, and System 2, which is rational and deliberative. The DSRL algorithm models this dual-process by dynamically adjusting the influence of each system based on the specific characteristics of the decision at hand, particularly the time delay until the reward and the magnitude of the reward. The key advantage of this approach is that it captures not only the traditional time discounting behavior observed in human decision-making but also the magnitude effect, where individuals tend to discount larger rewards less steeply than smaller ones.

The DSRL algorithm begins by initializing key system parameters, such as the weights associated with System 1 and System 2. These initial parameters reflect the baseline influence of each system before any specific decision-making context is considered. The algorithm then processes input information regarding the delay until the reward and the magnitude of the reward. Based on this input, the algorithm computes a weighted decision, where the relative contributions of System 1 and System 2 are adjusted dynamically. For decisions involving small, immediate rewards, System 1 exerts more influence, reflecting impulsive behavior. Conversely, for larger, delayed rewards, System 2 dominates, reflecting a more rational, long-term perspective.

One of the core features of the DSRL algorithm is its ability to learn from experience. As decisions are made and feedback is received regarding the outcomes of these decisions, the algorithm updates the weights assigned to each system. Over time, this reinforcement learning process allows the algorithm to improve its decision-making by fine-tuning the balance between impulsive and rational behavior. The ultimate goal of the DSRL algorithm is to optimize decision-making by appropriately balancing short-term desires with long-term goals, and to do so in a way that adapts to the characteristics of the decision environment.

The overall structure of the DSRL algorithm is depicted in Fig. 1. This diagram shows how the algorithm integrates information about the reward magnitude and time delay, computes the system weights, and uses reinforcement learning to adjust these weights based on feedback from previous decisions.

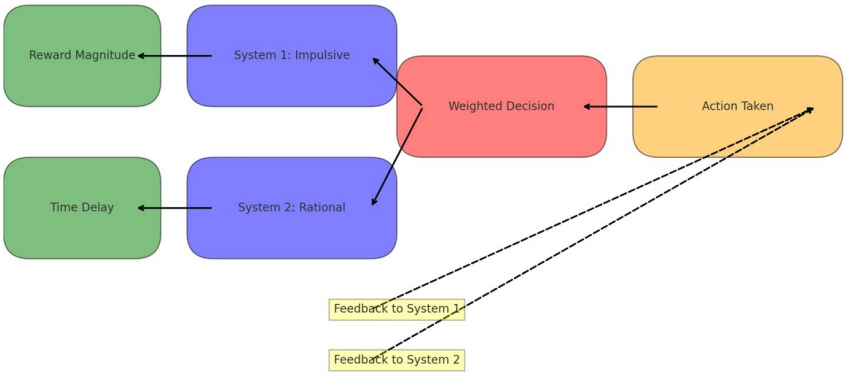

Fig. 1 System architecture of DSRL

2.2 Mathematical formulation

The DSRL algorithm is governed by a set of mathematical equations that define how the system weights are computed, how future rewards are discounted, how the value of a decision is calculated, and how the system updates itself over time. These equations form the backbone of the model and ensure that the algorithm behaves in a way that is consistent with both dual-system theory and reinforcement learning principles.

The first key component of the DSRL algorithm is the system weighting function, which determines how much influence System 1 and System 2 have on the decision-making process. This weighting function is dependent on two key factors: the time delay until the reward and the magnitude of the reward. The weighting function is defined as follows:

$$W(s_1, s_2) = \alpha(t)s_1 + \beta(R)s_2 \tag{1}$$

In this equation, $\alpha(t)$ is a function that depends on the time delay and controls the influence of System 1. As the delay increases, the value of $\alpha(t)$ decreases, reflecting the fact that impulsive decisions become less dominant when individuals are considering long-term rewards. On the other hand, $\beta(R)$ is a function of the reward magnitude, and it controls the influence of System 2. As the reward magnitude increases, $\beta(R)$ increases, reflecting the fact that larger rewards are more likely to be considered rationally, with System 2 dominating the decision-making process. This dynamic weighting of the two systems is a crucial aspect of the DSRL algorithm, as it allows the model to adapt to different decision contexts.

Next, the DSRL algorithm applies a discounting function to future rewards. This function reduces the value of rewards that will be received after a delay, reflecting the tendency of individuals to discount future outcomes. The discounting function used in the DSRL model is a hybrid of exponential and hyperbolic discounting, and it is defined as follows:

$$D(t, R) = \frac{1}{1 + k(t)t}(1 + \gamma \frac{R_0}{R}) \tag{2}$$

The first term of this equation captures the hyperbolic discounting behavior, where the discount rate decreases as the time delay increases. The second term captures the magnitude effect by reducing the discounting for larger rewards. R_0 represents a reference reward magnitude, and γ is a parameter that controls how sensitive the discounting process is to changes in reward size. This hybrid function allows the DSRL algorithm to accurately model both the time discounting and the magnitude effect.

To evaluate the expected outcome of a decision, the DSRL algorithm uses a value function. This function computes the expected reward from taking a particular action, combining both immediate rewards and discounted future rewards. The value function is given by the following equation:

$$V(s, a) = W(s_1, s_2)(r(a) + \gamma D(t, R)V(s', a')) \tag{3}$$

In this equation, $r(a)$ represents the immediate reward associated with action, and $V(s',a')$ represents the expected value of the next state after taking action. The discount function is applied to future rewards, ensuring that the model appropriately values both short-term and long-term outcomes.

The DSRL algorithm continuously improves its decision-making abilities through a learning rule that updates the system weights over time. This update rule is based on reinforcement learning principles and adjusts the weights based on the outcomes of previous decisions. The update rule is given by:

$$\theta_i^{t+1} = \theta_i^t + \eta(r(a) + \gamma V(s',a') - V(s,a))\nabla_{\theta_i} V(s,a) \tag{4}$$

Here, η represents the learning rate, which controls how quickly the model adapts to new experiences. The gradient term $\nabla_{\theta_i} V(s,a)$ reflects how sensitive the value function is to changes in the system weights. This learning rule ensures that the DSRL algorithm continually refines its decision-making process based on feedback from the environment.

Finally, the DSRL algorithm selects actions using an epsilon-greedy policy, which balances exploration (trying new actions) and exploitation (choosing the best-known action). This policy encourages the algorithm to occasionally explore new actions, while primarily selecting the action that maximizes the expected value based on the current state.

The effect of the DSRL algorithm's discounting process is depicted in Fig. 2, which illustrates how the model discounts future rewards based on both the time delay and the reward magnitude.

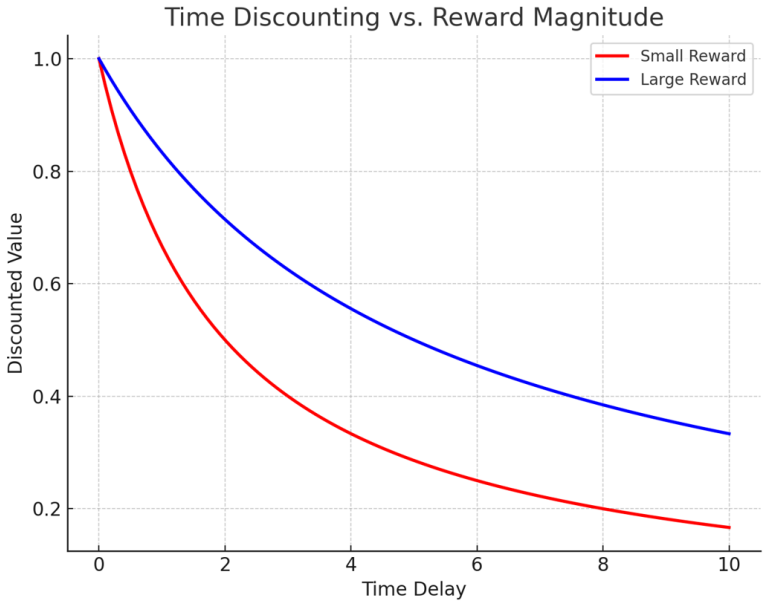

Fig. 2 Time discounting vs reward magnitude

2.3 Method diagrams

The reinforcement learning process within the DSRL algorithm plays a crucial role in refining decision-making over time by continuously adjusting the system weights for System 1 and System 2. After each decision, the model uses feedback on the reward received to update these weights, improving its ability to balance impulsive and rational decision-making in future scenarios. This feedback loop ensures that the model becomes progressively better at adapting to different decision contexts, such as varying reward magnitudes and time delays.

In each decision-making instance, the algorithm computes an output based on the current system weights. Once the outcome of the decision is observed, the DSRL algorithm evaluates the decision's quality by comparing the expected and actual rewards. If the reward exceeds expectations, the system responsible for the decision is reinforced by increasing its weight. Conversely, if the reward falls short of expectations, the corresponding system's influence is

reduced. This continuous update process allows the algorithm to shift the balance between impulsive and rational decision-making, enabling it to learn from both successes and failures.

The exploration-exploitation trade-off is managed using an epsilon-greedy policy, which allows the model to explore alternative actions while favoring those with the highest expected value. By periodically exploring new decision paths, the model avoids becoming trapped in suboptimal strategies and gains additional information that can lead to better long-term decisions. Over time, the algorithm relies more on exploitation, choosing actions with the highest expected return based on its learned experiences. Fig. 3 illustrates this learning process, showing how the system weights for System 1 and System 2 are updated after each decision. The feedback loop allows the model to refine its decision-making by adjusting the influence of each system, ultimately leading to improved performance across a range of decision-making tasks.

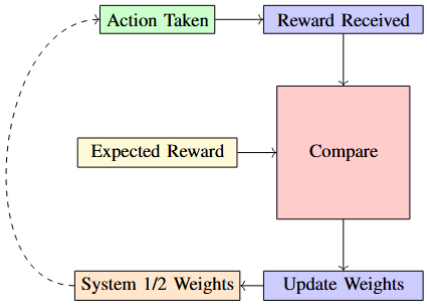

Fig. 3 Reinforcement learning process

As shown in the figure, the model's adaptability stems from its ability to continuously learn from past decisions. Early in the process, the model may experiment with different strategies through exploration, which enables it to gather valuable information. Over time, as it accumulates more experience, it becomes more proficient at predicting the outcomes of its actions, allowing it to make more rational and consistent decisions.

This feedback loop is essential for dynamically balancing the influence of impulsive and rational decision-making processes, based on the specifics of each decision context. By leveraging reinforcement learning, the DSRL algorithm effectively optimizes its decision-making strategy, ensuring that the system weights evolve in response to the rewards obtained. This adaptability is key to handling complex decision environments, such as those involving time delays and varying reward magnitudes.

Ultimately, the reinforcement learning mechanism ensures that the DSRL model continuously refines its strategy by learning from feedback. The balance between System 1 and System 2 evolves over time, leading to more informed and balanced decisions that reflect both short-term impulses and long-term rationality. The feedback loop, depicted in Fig. 3, is central to the model's ability to adjust dynamically, resulting in improved performance across a wide range of decision-making tasks.

3. Results and discussion

3.1 Experiment setup

In order to evaluate the performance of the DSRL algorithm, a series of decision-making experiments was conducted. These experiments simulate various decision scenarios where both the delay before rewards are received and the magnitude of rewards play a significant role in the decision-making process. The experimental setup includes a synthetic dataset that generates decision scenarios with varying reward sizes and time delays, reflecting real-world decision contexts.

The experiments were run with a fixed learning rate of 0.01 and an epsilon value of 0.1 for the epsilon-greedy policy. The reward sizes ranged from 10 to 100, and time delays were uniformly distributed between 0 and 10 units. Each experiment was repeated 10 times, and the results were

averaged to ensure statistical significance. Table 1 summarizes the key parameters used in the experiments, highlighting the range of variables that affect the decision outcomes.

The experimental setup is designed to capture the critical aspects of decision-making in the presence of time delays and varying reward magnitudes. By incorporating a wide range of reward sizes and delays, we ensure that the experiments provide a comprehensive evaluation of the DSRL model's ability to balance short-term impulsive actions with long-term rational decisions.

Table 1 Experimental setup parameters

Parameter	Description	Value
Learning rate	Controls how quickly the model learns	0.01
Epsilon	Controls exploration-exploitation trade	0.1
Reward Range	The range of rewards in the decision tasks	10-100
Time Delay	The time delay before rewards are received	0-10 units

3.2 Performance evaluation on time discounting

The first experiment focuses on the DSRL model's performance in handling time discounting, comparing it against traditional models such as exponential and hyperbolic discounting. Time discounting refers to the tendency of individuals to prefer smaller, immediate rewards over larger, delayed rewards. In this experiment, we evaluate how well the DSRL algorithm adjusts its decision-making based on the time delay before a reward is received.

Fig. 4 illustrates the performance of the DSRL model compared to baseline models. The graph shows how DSRL consistently outperforms the other models, particularly in decision contexts involving longer delays. This is because the DSRL model dynamically adjusts the weight of rational decision-making (System 2) as the time delay increases, allowing for more rational decisions when delays are significant. In contrast, the baseline models apply fixed discounting rates, leading to suboptimal decisions when delays are long.

The DSRL model accumulates higher rewards over time by favoring delayed but larger rewards when appropriate. The traditional models, on the other hand, tend to overemphasize immediate rewards, leading to poorer long-term outcomes. This experiment demonstrates that the DSRL model effectively captures the complexities of time discounting by balancing impulsive and rational decisions.

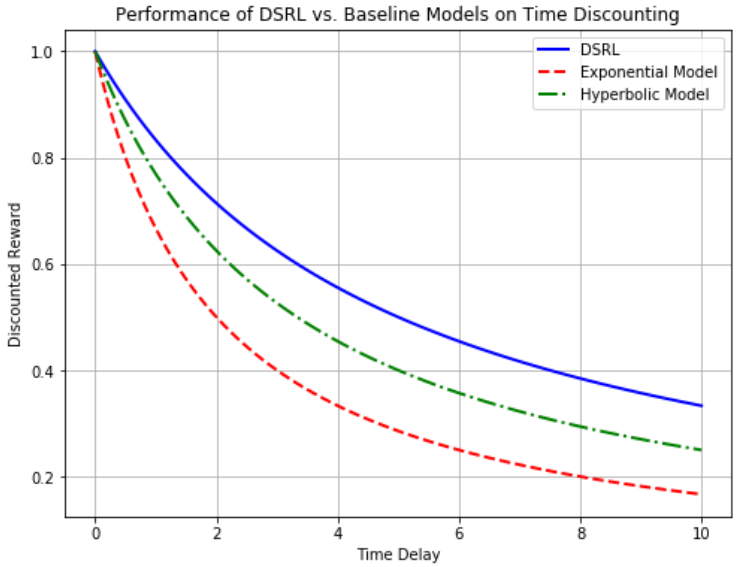

Fig. 4 Performance of DSRL vs baseline models on time discounting

3.3 Influence of reward magnitude

The second experiment examines the influence of reward magnitude on the DSRL algorithm's decision-making. It is well established in decision-making research that larger rewards are

discounted less steeply than smaller rewards, a phenomenon known as the magnitude effect. In this experiment, we assess how well the DSRL model adapts to varying reward sizes and adjusts its discounting behavior accordingly.

Table 2 presents the discount rates applied by the DSRL model and the baseline models for different reward sizes. The DSRL model applies lower discount rates to larger rewards, reflecting a more rational approach to evaluating high-magnitude outcomes. For smaller rewards, the DSRL model applies a higher discount rate, aligning with the tendency to prioritize immediate, smaller rewards. In contrast, the baseline models apply fixed discount rates across all reward sizes, failing to account for the difference in magnitude.

Fig. 5 further illustrates how the DSRL model adjusts its discounting based on reward size. The graph shows that larger rewards are discounted less steeply, resulting in higher long-term rewards. This ability to adapt to different reward magnitudes gives the DSRL model a significant advantage in decision scenarios where reward size plays a crucial role.

Table 2 Discount rates for different reward magnitudes

Reward size	DSRL discount rate	Baseline discount rate
Small (10-30)	0.15	0.25
Medium (40-70)	0.10	0.20
Large (80-100)	0.05	0.15

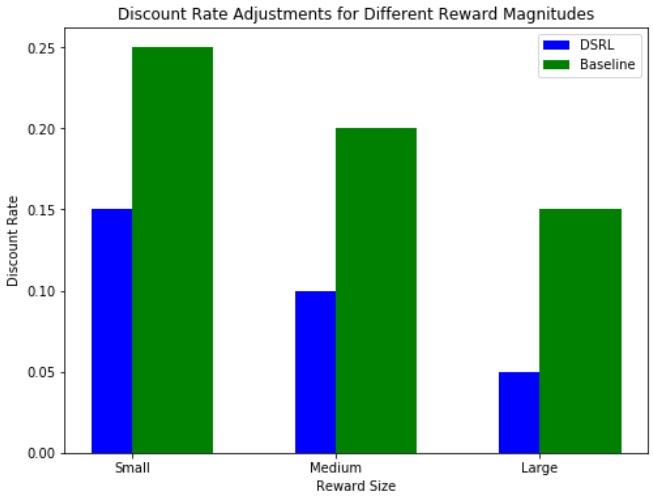

Fig. 5 Discount rate adjustments for different reward magnitudes

The results clearly show that the DSRL model performs better than the baselines by applying rational discounting that accounts for reward magnitude. This adaptability is crucial in real-world decision-making scenarios, where larger rewards are often more important than smaller, immediate ones.

3.4 Reinforcement learning adaptation

The final experiment evaluates how quickly and effectively the DSRL algorithm learns from feedback and improves its decision-making over time. We track the model's performance over multiple iterations to assess how quickly it converges to an optimal decision-making strategy. This experiment also compares the learning efficiency of the DSRL model to that of traditional models like exponential and hyperbolic discounting.

Table 3 summarizes the number of iterations required for each model to converge to an optimal decision-making strategy. As shown, the DSRL algorithm converges significantly faster than the baseline models, requiring fewer iterations to reach a high level of performance. Additionally, the DSRL model achieves a higher average reward, reflecting its ability to learn from experience and adjust system weights between impulsive and rational decision-making.

Fig. 6 shows the learning curves for the DSRL model and the baseline models. The DSRL model not only converges faster but also achieves higher rewards over time, demonstrating its superior learning capability. The ability to continuously adjust system weights through reinforcement learning allows the DSRL model to balance short-term impulsivity and long-term rationality more effectively than the baselines.

Table 3 Learning efficiency: DSRL vs baseline models

Model	Iterations to Convergence	Average Reward
DSRL	1000	85
Exponential model	2000	75
Hyperbolic model	1500	78

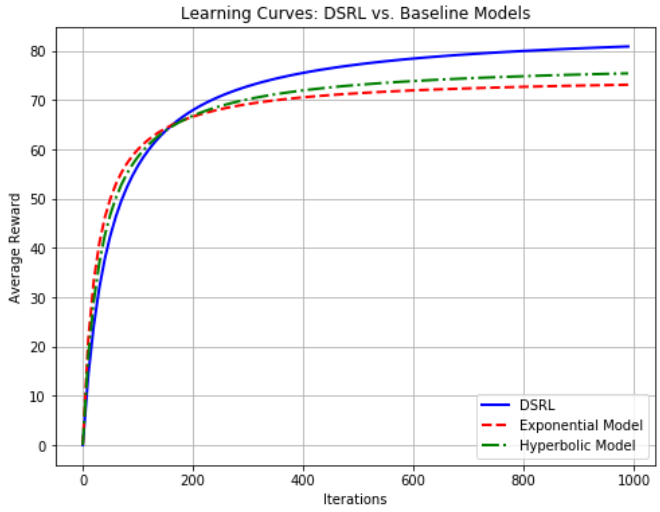

Fig. 6 Learning curves: DSRL v baseline models

The results from this experiment demonstrate the DSRL model's efficiency in learning and adapting to various decision-making contexts. Its ability to quickly converge to an optimal strategy, while maintaining high average rewards, highlights the robustness of the algorithm in balancing different types of decision behavior over time.

Through a series of experiments, we have demonstrated that the DSRL algorithm outperforms traditional models in various decision-making tasks. The model effectively captures both time discounting and the magnitude effect, and it shows superior learning efficiency in adapting to feedback. The ability to dynamically adjust the balance between impulsive and rational decision-making gives the DSRL model a significant advantage in handling complex decision scenarios, resulting in improved long-term performance.

4. Conclusion

This paper introduces the Dual-System Reinforcement Learning (DSRL) algorithm, which combines dual-system theory with reinforcement learning to model intertemporal decision-making. DSRL dynamically adjusts the balance between impulsive and rational decision-making systems based on reward magnitude and time delay, addressing key limitations in traditional discounting models like exponential and hyperbolic discounting. By doing so, DSRL more accurately captures how individuals weigh short-term versus long-term rewards. Experimental results confirm that DSRL outperforms baseline models, achieving higher cumulative rewards and greater learning efficiency. This demonstrates the model's robustness in handling complex environments where decisions involve delayed outcomes. Looking ahead, several directions for future research could further enhance the DSRL framework. Incorporating unsupervised learning techniques would improve the model's adaptability in uncertain or evolving reward structures, making it more flexible in diverse decision contexts. Additionally, integrating real-time decision-making capabilities would

enable DSRL to be applied to time-sensitive tasks such as autonomous systems or real-time financial trading, expanding its practical utility. Another promising area for future work is exploring how DSRL could be extended to multi-agent systems, where multiple decision-makers interact. This could provide insights into group dynamics and cooperation.

References

[1] Thaler, R. “Some empirical evidence on dynamic inconsistency”. Economics letters, vol. 8, no. 3, pp. 201-207, 1981.

[2] Frederick, S., Loewenstein, G., O’donoghue, T. “Time discounting and time preference: A critical review”. Journal of economic literature, vol. 40, no. 2, pp. 351-401, 2022.

[3] Frank, C. C., Seaman, K. L. “Aging, uncertainty, and decision making—A review”. Cognitive, Affective, & Behavioral Neuroscience, vol. 23, no. 3, pp. 773-787, 2023.

[4] Meissner, T., Gassmann, X., Faure, C., et al. “Individual characteristics associated with risk and time preferences: A multi country representative survey”. Journal of Risk and Uncertainty, vol. 66, no. 1, pp. 77-107,2023.

[5] Fiévet, A. “Decision over Time as a By-Product of a Measure of Utility: A Reappraisal of Paul Samuelson’s A Note on Measurement of Utility (1937)”. The European Journal of the History of Economic Thought, vol. 29, no. 3, pp. 438-454, 2022.

[6] Alcocer, C. D., Ortegón, J., Roa, A. “Uncertainty under hyperbolic discounting: the cost of untying your hands”. Journal of Economics, Finance and Administrative Science, vol. 24, no. 48, pp. 176-193, 2019.

[7] Schröder, D., Gilboa Freedman, G. “Decision making under uncertainty: the relation between economic preferences and psychological personality traits”. Theory and Decision, vol. 89, no. 1, pp. 61-83, 2020.

[8] Myers, P. S. “Anomalies in Intertemporal Choice: Evidence and a Interpretation”. Journal of Risk & Insurance, vol. 60, no. 1, pp. 1-10, 1993.

[9] Whitman, D. G., Rizzo, M. J. “The problematic welfare standards of behavioral paternalism”. Review of Philosophy and Psychology, vol. 6, pp. 409-425, 2015.