Learning Dynamics in Anxiety: The Role of Punishment Sensitivity and Learning Rate in Sequential Evaluation

Alicia Muñiz-Jiménez (ali.psi.neuro@gmail.com)
Department of Psychology, National Autonomous University of Mexico (UNAM), Mexico City, Mexico.

Víctor Mijangos (vmijangosc@ciencias.unam.mx)
Department of Science, National Autonomous University of Mexico (UNAM), Mexico City, Mexico.

Arturo Bouzas (arbouria@unam.mx)
Department of Psychology, National Autonomous University of Mexico (UNAM), Mexico City, Mexico.

Abstract

Anxiety is fundamentally an anticipatory response to uncertain future threats, making sequential evaluation crucial. However, most computational research has involved one-step tasks that identify an elevated punishment learning rate as a key feature of anxious behavior, while some studies also suggest a lack of evidence for the role of the punishment sensitivity parameter. This contradicts recent research involving sequential evaluation tasks, which supports the idea that anxiety is primarily an uncertainty disorder. Nevertheless, these experiments have mainly focused on model-based algorithms and have left the role of the punishment learning rate unexplored. To reconcile the differences in the literature and better understand how these two parameters influence anxiety, two hybrid models were developed: one with differentiated learning rates for rewards and punishments, and another with differentiated sensitivity parameters. The models were then evaluated in the Cliff Walking task, using a deterministic environment and incorporating stochasticity in action selection to simulate that agents did not have complete control over their decisions. The results suggest that the impact of estimated punishments on planning is more significant than the speed at which they are learned, highlighting the importance of heightened punishment sensitivity in anxiety-related behaviors such as avoidance, risk aversion, threat overestimation, and fear generalization.

Keywords: computational psychiatry; anxiety; reinforcement learning; sequential evaluation

Introduction

Computational modeling for studying anxiety is of great importance, as anxiety disorders have the highest incidence worldwide among mental disorders (Stein, Scott, de Jonge, & Kessler, 2017). Specifically, the reinforcement learning (RL) approach has aimed to construct mechanistic models that explain the underlying components and their interactions involved in anxious decision-making (Raymond, Steele, & Seriès, 2017; Pike & Robinson, 2022), where an altered part of the model describes anxious behavior (Huys, 2014), for example, a higher value in a certain parameter. However, there are mixed results regarding whether a higher punishment learning rate or an enhanced punishment sensitivity parameter plays a key role in anxious decision-making. This discrepancy becomes even more relevant when considering that these different conclusions stem from distinct kinds of tasks and, consequently, different types of models.

On one hand, most anxiety computational research has focused on simplified one-step tasks, such as multi-armed bandits, where an agent performs many trials but each trial ends after a single action, meaning the agent only needs to learn the appropriate response for a single environmental state (Yamamori & Robinson, 2023). Additionally, a systematic review and meta-analysis of 27 articles by Pike and Robinson (2022) supports that in studies using one-step tasks, an enhanced punishment learning rate is the primary factor for modeling anxious decision-making, while also indicating a lack of evidence for the role of the punishment sensitivity. Still, due to the nature of these tasks, planning has not been considered, and the models developed only try to replicate direct learning in one-state environments. Nevertheless, anxiety is an anticipatory response to uncertain future threats perceived as uncontrollable and highly aversive (Clark & Beck, 2010), involving a multi-step process where how planning may be biased should be considered (Sharp, 2025).

On the other hand, anxiety computational research involving sequential evaluation tasks is more recent and as a result comprises only a few studies, which primarily rely on variations of the value iteration algorithm to model biased offline planning and reproduce common behaviors observed in many anxiety disorders, such as avoidance, risk aversion, threat overestimation, and fear generalization (Zorowitz, Momennejad, & Daw, 2020; Gagne & Dayan, 2022). These works have found that heightened punishment sensitivity is a key feature to reproduce anxious behavior, highlighting that anxiety is mainly a disorder of uncertainty (Brown, Price, & Dombrovski, 2023), although the role of the punishment learning rate remains unexplored in multi-step tasks.

Therefore, while earlier studies on modeling anxious behavior emphasize the importance of how quickly punishment information is integrated over time, what matters in sequential evaluation experiments is how much an agent anticipates and dislikes being punished. As computational research on anxiety is relatively recent, no study has tested the contributions of both parameters —punishment learning rate and punishment sensitivity— simultaneously in a sequential evaluation task, which is a more natural setting for anxiety. To address this, it is necessary to use hybrid models that retain the core components of both approaches: direct learning in the environment and offline planning. Ultimately, this will help reconcile differences in the literature and offer deeper insights into the learning dynamics of anxiety in sequential evaluation tasks.

Among hybrid RL models, empirical evidence suggests that a hybrid successor representation (SR) is the most effective algorithm for replicating human decision-making in sequential evaluation tasks (Momennejad et al., 2017; Russek, Momennejad, Botvinick, Gershman, & Daw, 2017). Specifically, SR-Dyna is considered the best model for capturing human retrospective revaluation behavior (Momennejad, 2020). In addition, in SR family models rewards and successors (trajectories) are learned independently, and both can then be used to compute Q-values for each state (Ducarouge & Sigaud, 2017). Moreover, online planning is employed in the basic SR algorithm when successors are estimated or used to compute the estimation of another variable (Gershman, 2018). Therefore, the SR-Dyna model implements direct experience within the environment, online planning, and offline planning. The integration of these three mechanisms — with some modifications — may better reflect how individuals with anxiety process information in real-world situations, compared to models that incorporate only one or two of them. For example, for a woman with acrophobia, direct learning instantly increases her association of heights with threat after falling from a cliff. Meanwhile, offline planning involves rumination on various possible situations based on past experiences while at rest. Finally, online planning manifests as hypervigilance for potential future threats when she's actually at heights and has to consider trajectory estimations, with all three mechanisms reinforcing her fear of heights.

Furthermore, the independence in learning between rewards and successors in SR models allows us to make specific modifications to the model's equations to assess the impact of differentiated learning rates and sensitivity parameters, considering distinct learning dynamics. For example, a first model can be constructed by implementing differentiated learning rates by modifying the reward function. In contrast, a second model can be designed with differentiated sensitivity parameters for rewards and punishments in the successor function. Next, simulations using different values of these parameters can assess the impact on anxious behavior of a higher punishment learning rate in the first model and an enhanced punishment sensitivity parameter in the second one. Hence, we propose using an SR-Dyna algorithm as a basis to design new models to address the objective of this work.

Methodology

The present work adopts a theoretical modeling approach to generate predictive insights for anxiety research. Two models were developed to simulate agents with varying levels of punishment learning rate and punishment sensitivity, with the aim of evaluating their respective contributions to anxious decision-making within a Markov Decision Process (MDP).

Models

In algorithm 1 we present the pseudocode for an SR-Dyna model considering a deterministic environment, incorporating temporal difference (TD) learning and estimating Q-values, followed by a detailed explanation of the modifications introduced to it for the first and second models proposed in this research.

Algorithm 1 : SR-Dyna

1: **Initialize** $R(s)\ \forall\ s \in S$.
2: **Initialize** $M(s,s',a)\ \forall\ s,s' \in S$ and $a \in A(s)$.
3: **Initialize** $Q(s,a)\ \forall\ s \in S$ and $a \in A(s)$.
4: **Initialize** $E(s,a)\ \forall\ s \in S$ and $a \in A(s)$.
5: **while** episode $\leq N$ **do**
6: $s \leftarrow$ actual state
7: $a \leftarrow$ ε-greddy(s,Q)
8: Execute action a; observe the received reward (r), and the next state (s').
9: $a^+ \leftarrow \arg\max_a Q(s',a')$
10: $M(s,\mathbf{s}',a) \leftarrow M(s,\mathbf{s}',a) + \alpha(I_{[s=s']} + \gamma M(s',\mathbf{s}'',a^+) - M(s,\mathbf{s}',a))$
11: $R(s) \leftarrow R(s) + \alpha(r - R(s))$
12: $Q(s,a) \leftarrow \sum_{s'} M(s,\mathbf{s}',a)R(s')$
13: $E(s,a) \leftarrow r,s'$
14: **for** n **do**
15: $s \leftarrow$ previously observed random state.
16: $a \leftarrow$ random action previously executed in s.
17: $r,s' \leftarrow E(s,a)$
18: Execute steps from **9** to **12** using the new values for s, a, r and s'.
19: **end for**
20: **end while**

Parameters: N, ε, α, γ and n.

First, in Algorithm 1 the vector for rewards (R), the successor tensor (M), the Q-value matrix (Q) and the matrix with the environment's model (E) are initialized with zeros in each entry. Then, during each episode in the model-free part of the model, the agent selects an action (a) using an ε-greedy policy, executes it, receives a reward (r) and observes the next state (s'). Later, the agent updates the current successor representation using the TD method, incorporating both immediate and discounted future occupancy predictions. Additionally, the estimated reward, Q-value estimates, and the environment's model are updated. In the Dyna-style offline planning phase (steps 14-18), the algorithm replays stored experiences using the environment's model and updates the SR model estimations.

Model 1: Dyna α-SR This model differs from Algorithm 1 solely in how the learning rate (α) is selected within the reward function (step 11). Specifically, α value now depends on whether the agent receives a reward or a punishment, as shown in Equation 1, in order to have differentiated learning rates.

$$\alpha = \begin{cases} \alpha^+ & \text{if } r \geq 0 \\ \alpha^- & \text{if } r < 0 \end{cases} \tag{1}$$

Model 2: Dyna β-SR Building on Algorithm 1 and the work of Zorowitz et al. (2020) to model anxious behavior, we introduce a novel variant: Dyna β-SR. In this new model, it is necessary not only to calculate the action with the highest value (a^+) for the next state, but also the lowest (a^-), as shown in Equations 2 and 3.

$$a^{+} \leftarrow \arg\max_{a} Q(s', a') \tag{2}$$

$$a^{-} \leftarrow \arg\min_{a} Q(s', a') \tag{3}$$

Later, a^+ and a^- are used in the β^M module (Eq. 4) to estimate the successors with the highest and lowest action for the next state, which are multiplied by ω and its complement, respectively. Therefore, we can say that a pessimistic agent (ω = 0) is sensitive to the paths that lead to punishments or threats, considering that it lacks complete control over its actions. This can be considered as a low level of self-efficacy, reflecting a lack of belief in their ability to perform the necessary behaviors to achieve specific performance goals (Bandura, 1978; Zorowitz et al., 2020). In contrast, an optimistic agent (ω = 1) anticipates that its future actions will fully align with its preferences to maximize rewards (Zorowitz et al., 2020), being sensitive to the paths that lead to them. Finally, β^M is used in the successor function (Eq. 5) as the successor representation for the next state.

$$\beta^{M} \leftarrow \omega M(s', \mathbf{s}'', a^{+}) + (1-\omega) M(s', \mathbf{s}'', a^{-}) \tag{4}$$

$$M(s, \mathbf{s}', a) \leftarrow M(s, \mathbf{s}', a) + \alpha \left(I_{[s=s']} + \gamma \beta^{M} - M(s, \mathbf{s}', a) \right) \tag{5}$$

Experimental Task

A variation of the Cliff Walking task (Fig. 1) was implemented in a 9x9 gridworld with a discrete-time environment, finite horizon, and deterministic dynamics. The stochasticity lies in the agent's action selection, as it follows an ε-greedy policy to choose the action it will execute in each state. This policy is used both during the agent's training phase ($N = 200$ episodes) and in the final test (1 rollout) to simulate that the agent does not have complete control over its actions, as is the case in many real-world scenarios.

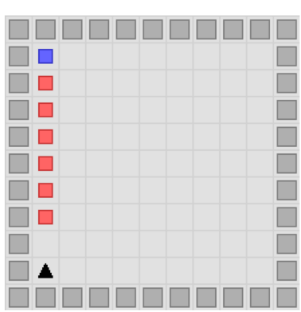

Figure 1: **Virtual Environment of the Cliff-Walking task.** The triangle indicates the agent's starting position, the blue square marks the goal, the red squares represent aversive states (the cliff), and the light gray squares are neutral, where neither a reward nor a punishment is received.

The agent's objective in this environment is to reach the goal without falling into an aversive state. However, at the beginning, the agent is unfamiliar with the environment and will need to explore it to find the optimal route. The agent can perform four possible actions: moving up, down, right, or left, as long as it does not collide with a wall (dark gray squares). Finally, an episode can end under the following conditions: the agent reaches the goal (reward = 1), the agent falls into an aversive state (punishment = -1), or the maximum step limit per episode is reached (100).

Additionally, the anxious behaviors we aimed to evaluate in the experimental task were operationalized as follows:

- **Threat overestimation:** states that would otherwise be neutral are estimated as signaling threat unrealistically.
- **Fear generalization:** neutral states that are distant from aversive ones are considered dangerous.
- **Avoidance:** the agent consistently moves away from aversive states.
- **Risk aversion:** instead of exploring more of the environment, the agent exploits a safe zone of the state space.

Simulations

Three simulations were conducted for the Dyna α-SR agents with α^- values of 0.05, 0.075, and 0.1, respectively, while keeping $\alpha^+ = 0.05$ constant in all cases. In addition, three simulations were conducted for the Dyna β-SR agents with ω values of 1.0, 0.5, and 0.0, respectively. The parameters shared between the models were set as follows: α = 0.1 (for the successor function), $\gamma = 0.9$, ε = 0.2 and n = 3 (recall). All simulations were performed using the NeuroNav library (Juliani, Barnett, Davis, Sereno, & Momennejad, 2022) as a foundation for programming the agents and the experimental task. The complete code is publicly available at the following link: `https://github.com/Alicia-MJ/RL-Anxiety-MathPsy-ICCC.git`.

Results

For each agent, a heatmap of the environment was generated based on the final estimated value of each state, computed by averaging its corresponding Q-values. In these maps, values close to 1.0 (the maximum possible value) appear in deep blue, values near 0 are shown in white, and values close to -1.0 (the minimum possible value) are displayed in dark red. Arrows indicate the highest-value action estimated for each state, while the rollout corresponding to the evaluation phase is highlighted with black arrows. Additionally, as a support to the main analysis, the total number of steps and the total return received per episode were graphed for each agent to evaluate its evolution across episodes. Together, these three analyses provide a more comprehensive assessment of each agent's behavior and learning.

Dyna α-SR Agents

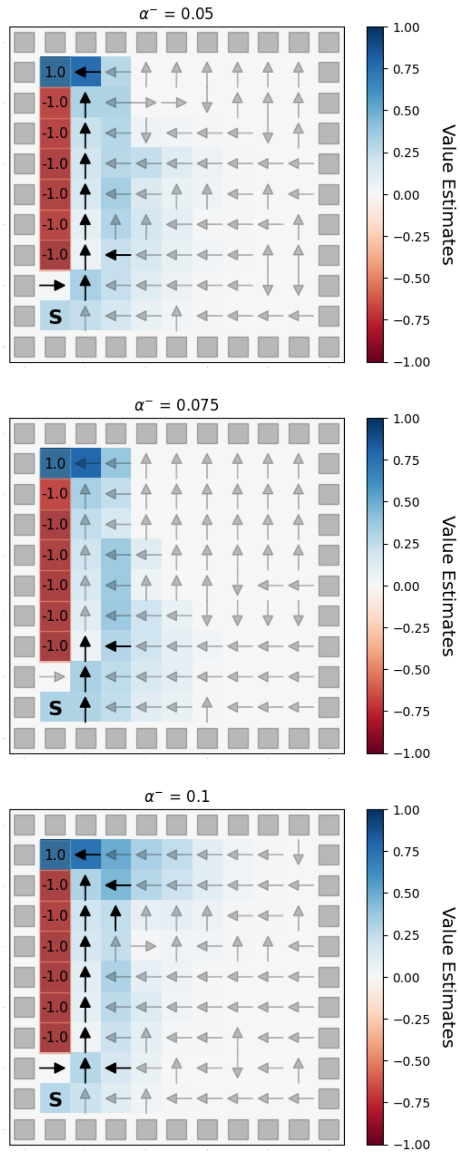

Figure 2: **Dyna α-SR agents' environment estimation.** The top image corresponds to the agent with $\alpha^- = 0.05$ (reward and punishment learning rates are equal), the middle to the agent with $\alpha^- = 0.075$ (50% higher than α^+), and the bottom to the agent with $\alpha^- = 0.1$ (double than α^+).

The three Dyna α-SR agents show very similar behavior across the three analyses conducted. First, Figure 2 demonstrates that their environmental estimations are highly similar, even in the case where an agent had a punishment learning rate (α^-) that was twice the value of the reward learning rate (α^+). Additionally, all three agents chose the shortest but unsafe path to the goal during the testing phase (indicated by black arrows), underestimating danger. While the first and third agents managed to reach the goal in the rollout, the second agent followed a similar path but, due to the stochasticity in action selection, ended up falling into the cliff.

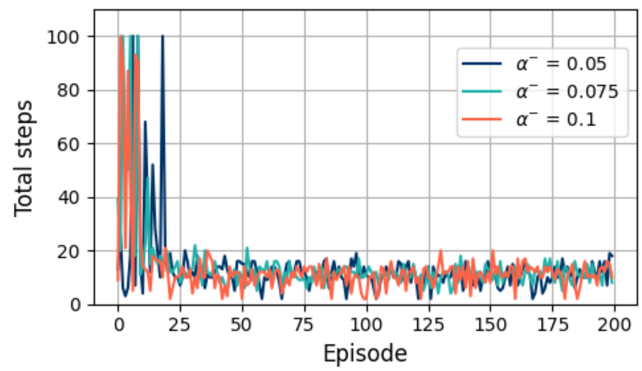

Figure 3: **Dyna α-SR agents' total steps per episode.**

Moreover, Figure 3 shows that the agents required more temporal steps at the beginning of the learning phase, which then decreased drastically and converged around similar values, consistent with their preference for the shortest path to the goal. Additionally, due to the similarities in both environmental estimation and the number of steps per episode, a similar level of goal achievement was expected. This is confirmed in Figure 4, where the agents consistently oscillated between reaching the goal (total return = 1) and falling into the cliff (total return = -1) across episodes, reflecting the inherent risk of choosing the shortest path. Few exceptions occurred at the beginning, when the agents sometimes remained in neutral states (total return = 0) during early learning. Finally, minor differences between agents can be attributed to the stochasticity of the programmed policy.

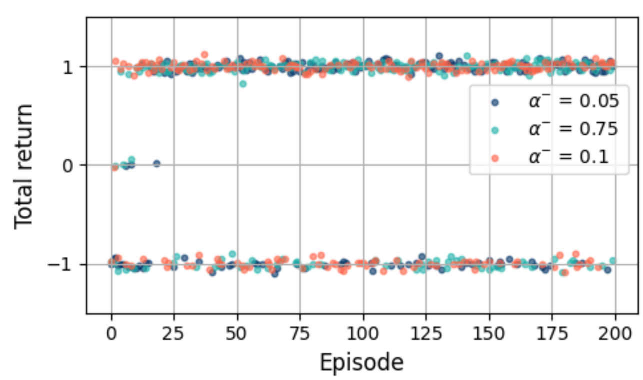

Figure 4: **Dyna α-SR agents' total return per episode.**

In summary, the results for the Dyna α-SR agents show that a higher punishment learning rate did not lead to the emergence of anxious behaviors such as threat overestimation, fear generalization, avoidance, or risk aversion, within the experimental task.

Dyna β-SR Agents

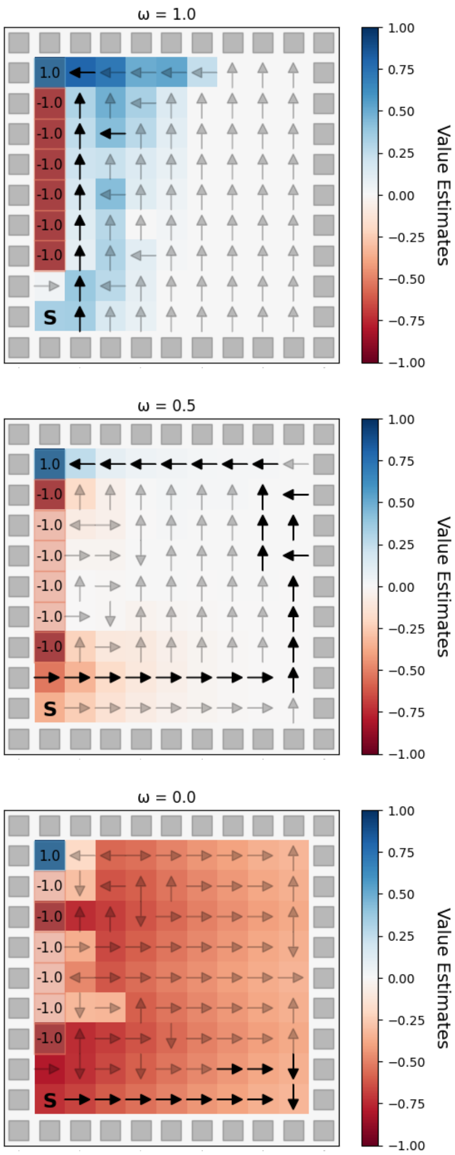

Figure 5: **Dyna β-SR agents' environment estimation**. The top image corresponds to the optimistic agent (ω = 1), the middle to the agent with ω = 0.5, and the bottom to the pessimistic agent (ω = 0).

In contrast to Dyna α-SR agents, Dyna β-SR agents show a marked difference in learning and behavior across the analyses conducted. To begin with, Figure 5 shows that the optimistic agent (ω = 1) learned to take a short but risky route to the goal near the cliff by estimating nearby states positively, without accounting for the possibility of not always being able to choose its own actions. Conversely, the moderately optimistic–pessimistic agent (ω = 0.5) avoided the cliff during the testing phase, as it evaluated nearby states as dangerous. Nevertheless, it was still able to reach the goal by taking a longer but safer path. Finally, the pessimistic agent (ω = 0) estimated much of the environment as aversive and tried to stay away from the cliff during the rollout, but ultimately failed to reach the goal.

Moreover, both the ω = 0.5 and ω = 0.0 agents exhibited avoidance behavior, but the extremely pessimistic agent was the one that excessively overestimated threat and demonstrated fear generalization. Additionally, the results suggest that threat overestimation and fear generalization bias reward evaluation, leading the agent to prioritize danger avoidance over goal pursuit. This reflects risk aversion, as the agent exploits a safe zone of the state space instead of exploring the environment further.

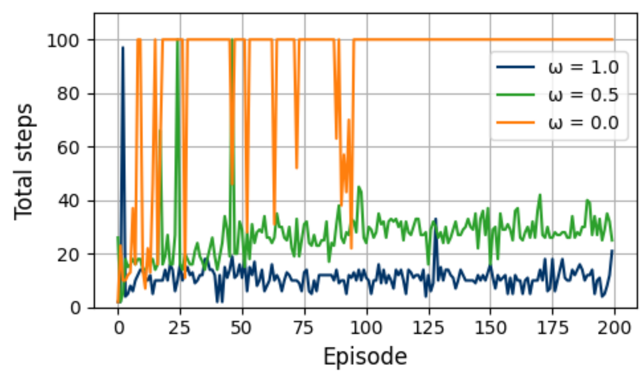

Figure 6: **Dyna β-SR agents' total steps per episode.**

Furthermore, the results from Figure 5 align with those in Figure 6, which show that the optimistic agent (blue) used the fewest temporal steps in most episodes by taking a more direct route to the goal. It was followed by the moderately optimistic–pessimistic agent (green), which took a longer but safer path to the goal. Meanwhile, the pessimistic agent (orange) used the maximum number of steps (100) in most episodes, as it tended to remain in neutral states far from danger

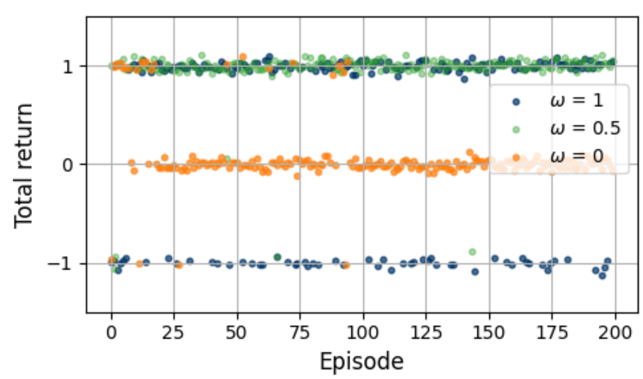

Figure 7: **Dyna β-SR agents' total return per episode.**

Lastly, the patterns observed for each agent in Figure 7 complement the previous results by highlighting how the agents' environment estimations and learned strategies impacted their actual outcomes. The optimistic agent (blue) alternated between reaching the goal and falling into the cliff. The agent with $\omega = 0.5$ (green) achieved the goal most consistently, ending in neutral or aversive states only occasionally. In contrast, the pessimistic agent (orange) remained in neutral states for most episodes, particularly in the last 100 episodes.

In summary, the results for the Dyna β-SR agents demonstrate that enhanced punishment sensitivity was directly linked to the exhibition of anxiety-related behaviors in the experimental task, as well as a level of punishment sensitivity being necessary for an agent to identify a safer path to the goal.

Discussion

In this work, we expanded anxiety computational research in sequential evaluation tasks by using hybrid reinforcement learning models to assess the influence of punishment learning rate and punishment sensitivity in order to better understand the learning dynamics underlying anxious behavior. The simulation results suggest that the impact of estimated punishments on planning is more significant than the speed at which they are learned, highlighting the importance for heightened punishment sensitivity in anxious decision-making.

Specifically, the behavioral similarity observed among Dyna α-SR agents indicates that a higher punishment learning rate does not promote the emergence of anxious decision-making in the experimental task. In contrast, the performance differences among the Dyna β-SR agents support the idea that increased sensitivity to punishment facilitates the emergence of anxiety-related behaviors, such as avoidance, risk aversion, threat overestimation, and fear generalization. The Dyna β-SR model, in particular, predicts that pessimistic planning —arising from an enhanced sensibility to threat due to the agent's underestimation of its own self-efficacy or coping resources— may play a central role in the development of these behaviors. This interpretation stems from the observation that what matters most is not how quickly punishments are learned, but how their estimations influence the updating of successor representations, which can lead to a progressive distortion of the cognitive map built by the agent when it has high sensitivity to punishments. In essence, the model highlights the role of catastrophic thinking or worry in anxiety, conceptualized here as a tendency to prioritize the most disadvantageous routes during both online and offline planning.

These findings align with previous research that considered sequential evaluation (Brown et al., 2023; Gagne & Dayan, 2022; Zorowitz et al., 2020), which holds anxiety disorders are mainly disorders of uncertainty where a heightened punishment sensitivity might be a key feature to predict anxious behavior. On the other hand, the results challenge aspects of prior research based on one-step tasks (Pike & Robinson, 2022), which have emphasized the role of rapid punishment integration in the emergence of anxious decision-making. This raises the question of whether such tasks are adequate for studying anxiety, as they might not fully capture its complexity and could bias our understanding. Alternatively, it could be considered that different task structures —one-step versus multi-step— help clarify the specific ways in which a high punishment learning rate contributes to anxiety. For example, a high learning rate for punishments may be more relevant in the short term, as observed in one-step tasks, while it might lose relevance in sequential evaluation, where a high punishment sensitivity parameter becomes much more significant. Nevertheless, given that anxiety is primarily an anticipatory response to uncertain future threats perceived as uncontrollable and highly aversive (Clark & Beck, 2010), findings from sequential evaluation tasks may provide a more accurate understanding of the factors influencing anxious behavior.

Finally, this work provides valuable insights into the learning dynamics of anxiety considering sequential evaluation, although future research should explore other experimental tasks where differentiated learning rates may play a more significant role, particularly in non-deterministic environments. To achieve this, it will be essential to extend the models' offline planning module to account for uncertainty in environmental dynamics and enhance their robustness. Additionally, it will be interesting to determine the number of replays an agent makes, which can be linked to the degree of pessimism or surprise. Ultimately, an important next step should be to evaluate the Dyna β-SR model's predictions against human data to assess its plausibility beyond simulation.

Acknowledgements

We thank Dr. Nathaniel Daw and his lab members for their valuable feedback on the research, as well as Dr. Carlos Velázquez, PhD candidate José Luis Baroja, and the three reviewers for their constructive feedback on the paper. Additionally, this work was supported by grants from the CONAHCYT project A1-S-11703, the PITAAE 2024 fellowship, and the DGAPA-PAPIIT projects BG101721 and TA100924.

References

Bandura, A. (1978). Self-efficacy: Toward a unifying theory of behavioral change. *Advances in Behaviour Research and Therapy*, *1*(4), 139-161.

Brown, V. M., Price, R., & Dombrovski, A. Y. (2023). Anxiety as a disorder of uncertainty: implications for understanding maladaptive anxiety, anxious avoidance, and exposure therapy. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(3), 844–868.

Clark, D. A., & Beck, A. T. (2010). *Cognitive therapy of anxiety disorders: Science and practice*. New York, NY: The Guilford Press.

Ducarouge, A., & Sigaud, O. (2017). The successor representation as a model of behavioural flexibility. In

Journées francophones sur la planification, la décision et l'apprentissage pour la conduite de systèmes. Caen, FR: HAL.

Gagne, C., & Dayan, P. (2022). Peril, prudence and planning as risk, avoidance and worry. *Journal of Mathematical Psychology*, *106*, 102617.

Gershman, S. J. (2018). The successor representation: Its computational logic and neural substrates. *Journal of Neuroscience*, *38*(33), 7193–7200.

Huys, Q. (2014). Computational psychiatry. In D. Jaeger & R. Jung (Eds.), *Encyclopedia of computational neuroscience.* New York, NY: Springer New York.

Juliani, A., Barnett, S., Davis, B., Sereno, M., & Momennejad, I. (2022). Neuro-nav: A library for neurally-plausible reinforcement learning. In *The 5th multidisciplinary conference on reinforcement learning and decision making.* Providence, RI.

Momennejad, I. (2020). Learning structures: Predictive representations, replay, and generalization. *Current Opinion in Behavioral Sciences*, *32*, 155–166.

Momennejad, I., Russek, E. M., Cheong, J. K., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, *1*(9), 680–692.

Pike, A. C., & Robinson, O. J. (2022). Reinforcement Learning in Patients With Mood and Anxiety Disorders vs Control Individuals: A Systematic Review and Meta-analysis. *JAMA Psychiatry*, *79*(4), 313–322.

Raymond, J. G., Steele, J. D., & Seriès, P. (2017). Modeling trait anxiety: From computational processes to personality. *Frontiers in Psychiatry*, *8*, 1.

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Computational Biology*, *13*(9), e1005768. doi: 10.1371/journal.pcbi.1005768

Sharp, P. B. (2025). Anxiety involves altered planning. *Trends in Cognitive Sciences*, *29*(2), 118-121.

Stein, D. J., Scott, K. M., de Jonge, P., & Kessler, R. C. (2017). Epidemiology of anxiety disorders: from surveys to nosology and back. *Dialogues in Clinical Neuroscience*, *19*(2), 127–136.

Yamamori, Y., & Robinson, O. J. (2023). Computational perspectives on human fear and anxiety. *Neuroscience and Biobehavioral Reviews*, *144*, 104959.

Zorowitz, S., Momennejad, I., & Daw, N. D. (2020). Anxiety, avoidance, and sequential evaluation. *Computational Psychiatry (Cambridge, Mass.)*, *4*, 10.1162/cpsy_a_00026.