# Perseverative Behavioral Sequences Aid Long-Term Credit Assignment

**Sienna Bruinsma (sienna_bruinsma@brown.edu)**
Carney Institute for Brain Science, Brown University, 164 Angell St, Providence, RI 02906, USA

**Frederike Petzschner* (frederike_petzschner@brown.edu)**
Carney Institute for Brain Science, Brown University, 164 Angell St, Providence, RI 02906, USA
Department of Psychiatry & Human Behavior, Alpert Medical School of Brown University,
222 Richmond St, Providence, RI 02903, USA

**Matthew Nassar* (matthew_nassar@brown.edu)**
Carney Institute for Brain Science, Brown University, 164 Angell St, Providence, RI 02906, USA

**co-senior authors*

## Abstract

**Learning from the past requires assigning credit to the consequences of our actions. Here, we explored human credit assignment strategies by asking people to select activities with short- and long-term pain-related consequences for an avatar and predict the avatar's subsequent pain level. Human behavioral results suggest that, while participants can learn short-term consequences, their learning of long-term consequences depends critically on how they sequence activities in time. More specifically, increased repetition in activity selection (i.e., perseveration) is related to a learned preference for activities that reduce long-term, but not short-term, pain. Additionally, in comparing several computational models, we found that standard model-free algorithms (i.e., temporal difference learning) best explained the behavior of participants who did not perseverate, whereas Bayesian inference models that take into account the causal structure of the environment better explained those that did. Our results demonstrate that credit assignment critically depends on the order in which actions are selected, with repetitions aiding the learning of long-term consequences. This raises the possibility of perseveration as a useful action policy to improve long-term credit assignment.**

**Keywords:** credit assignment; perseveration; pain; temporal difference learning; Bayesian causal inference

## Introduction & Methods

In order to update future behavior, it is essential to learn whether a past action or event led to a good or bad outcome and assign credit accordingly (Sutton, 1984). For instance, if you feel worse than usual today, it is worth determining which actions (e.g., insufficient sleep, not eating breakfast) may have contributed to that state in order to improve choices in the future. However, assigning credit is a notoriously difficult problem for both humans (Gureckis & Love, 2009; Colaizzi et al., 2020) and animals (Robinson & Flagel, 2009) alike, especially when we want to consider the longer-term consequences of our actions given the lack of a clear temporal linkage between action and state in this case.

In the present study, we investigated how people assign credit to activities with immediate versus lasting consequences. To do so, we utilized two paradigms – a "Prediction Task" and an "Assessment Task" (Fig. 1a). During the Prediction Task, online participants (n=300) must choose between 4 different activities for an avatar to engage in and then predict the avatar's "pain level" after engaging in the selected activity. Whereby activities can have either short- (S) or long-term (L) consequences on the avatar's pain level (increase or decrease; see activities matrix Fig 1c). Participants then receive feedback about the avatar's actual pain level, which is displayed alongside their previous prediction on the next trial to reduce memory load (Fig 1a).

Activities impact the avatar's pain either by directly affecting the observed reported pain level ($x$) or by affecting the underlying mean pain ($\mu$), which is stored across trials and affects $x$ on each (Fig 1b). If the current activity has a short-term effect, $S_{A(t)}$, it will directly impact $x$ on the current trial, such that $x_t = \mu_t + S_{A(t)}$. In contrast, long-term effects on pain, $L_{A(t)}$, are implemented via a change in the underlying mean pain level ($\mu$) on the next trial, such that $\mu_{t+1} = \mu_t + L_{A(t)}$ (Fig 1b). Thus, long-term consequences of activities have a delayed and persistent impact on pain, whereas short-term consequences are immediate and transient (Fig 1c). After participants learn the short- and long-term consequences of activities via trial-and-error during the Prediction Task, we test their learned preferences for these activities in a two-alternative choice Assessment Task where they are instructed to choose the activity that will most minimize pain (Fig 1a). No feedback was given in the Assessment Task.

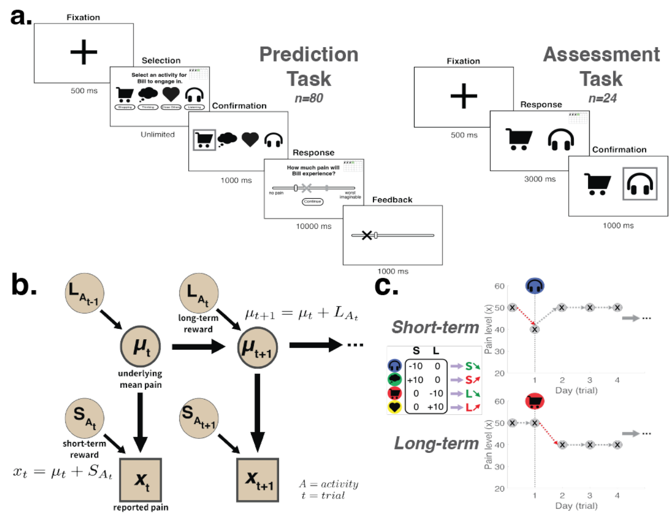


Figure 1: **a)** Task design **b)** Generative model for how pain is updated, with a sampling statement (bottom left) and transition function (top right) **c)** Left: activities matrix mapping activities to their associated effects; right: descriptive plots of short- and long-term effects of activities

## Behavioral Results

In a previous CCN paper (Bruinsma, Petzschner, & Nassar, 2023), we demonstrated how experimental manipulation of the ordering of activities in the Prediction Task had significant effects on participants' learning of short- and long-term consequences. Specifically, we showed that participants failed to learn long-term consequences unless activities were repeated multiple times in a row. In the present study, we gave participants the option to select activities (Fig 1a) to explore whether individual differences in activity selection patterns are directly related to improved learning of short- and long-term consequences.

In order to quantify participants' selection patterns, we utilized a metric called repeat frequency (RF). A participant with a low RF would be more likely to switch to a new activity, whereas a high RF participant would be more likely to repeat activities back-to-back. When correlating participants' RF in the Prediction Task (PT) with their performance in the Assessment Task (AT) (i.e., likelihood of choosing the pain-reducing short- or long-term activity) (Fig 2a), we find that RF explains much of the individual learning differences in our task. Specifically, greater repetition of activities (i.e., higher RF) resulted in a significantly better understanding of long-term consequences, while activity switching (i.e., lower RF) led to a worse understanding. We also identified a weaker but still significant opposite pattern for short-term activities, suggesting that there may be a trade-off between learning short- versus long-term consequences dependent on how individuals sequence activities in time (Fig 2a). Additionally, a trial-to-trial linear regression on the impact of different factors in their ability to predict an avatar's pain level in the PT shows a similar trend, where high RF individuals (as determined by a median split) are significantly more responsive to long-term activities than low RF individuals (Fig 2c). It is important to note, however, that participants tend to misattribute long-term effects to the current, rather than previous, activity (Fig 2c), suggesting that they are not fully identifying the causal structure and timing of long-term effects (see Fig 1b,c).

Lastly, in order to assess whether enhanced learning from different activity sequences is specific to those who actively explored the space or can generalize, we presented a yoked group of participants (n=150) with a predetermined sequence of activities matched to the original participant's sequences (n=150) and saw similar short- and long-term learning trends (Fig 2b,c).

## Computational Modeling

In order to understand the potential computations underlying participant behavior, we fit standard model-free algorithms (temporal difference (TD) learning with and without memory (TD-3)) and models accounting for the causal structure of the environment (Bayesian causal inference over possible activities matrices) to the participants' PT behavior. A Bayesian model selection based on the Bayesian Information Criterion showed that a 3-step TD model is best able to capture low RF participants, while high RF participants are better explained by a Bayesian inference model (Fig 2d).

Further analyses show that TD models are incapable of proper learning of long-term effects, while a Bayesian inference model shows optimal learning of short- and long-term effects but vastly outperforms participants. This makes sense given that long-term effects do not affect the state (formulated here as the activity) and TD only learns through future states. Thus, since long-term effects are not related to the future state (i.e., next activity), TD models struggle to learn them. In contrast, the Bayesian inference model has knowledge of the structure of the environment (i.e., the activities matrix; Fig 1C), although it must perform inference over hundreds of thousands of activities matrices to infer the correct values. Nonetheless, none of our models can fully explain participant behavior across different activity sequences and, thus, our next step is to develop variants of these models that better account for these unique constraints on human credit assignment.

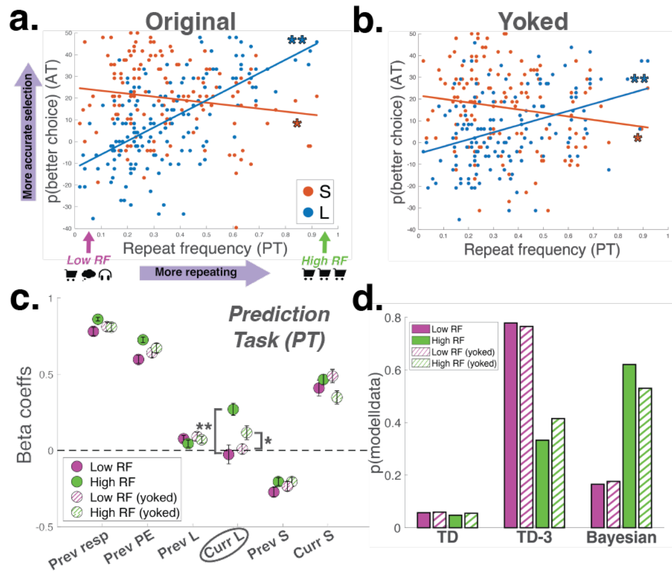


Figure 2: **a, b)** Correlation of RF in the prediction task (x-axis) with assessment task performance (y-axis) for short- (orange) and long-term (blue) activities, where dots are individual participants **c)** Linear regression on factors (x-axis) predicting participants' pain predictions in the PT (y-axis) **d)** Model likelihoods given by BMS

## Discussion

Overall, our results demonstrate that 1) there is a high amount of variability between individuals in how they explore and sequence their actions, and 2) that these sequencing differences lead to major differences in how people learn and assign credit, where greater perseveration (i.e., action repetition) aids the learning of long-term consequences. While perseveration is often regarded as a faulty policy utilizing a minimally complex model, our results challenge this notion and suggest that perseveration might actually be a useful form of structured exploration that allows participants to get a handle on the long-term consequences of actions. We intend to explore the real-world implications of these results, specifically in the domain of pain where we will assess whether the ability to learn the long-term consequences of activities on pain can predict the risk of individuals to develop chronic pain.

## Acknowledgements

This work was supported by the Brainstorm Program at the Robert J. & Nancy D. Carney Institute for Brain Science.

## References

Bruinsma, S., Petzschner, F., & Nassar, M. (2023). Long-Term Credit Assignment in Humans Critically Depends on Sequential Structuring of Events. *2023 Conference on Cognitive Computational Neuroscience.* Oxford, UK.

Colaizzi, J. M., Flagel, S. B., Joyner, M. A., Gearhardt, A. N., Stewart, J. L., & Paulus, M. P. (2020). Mapping sign-tracking and goal-tracking onto human behaviors. *Neuroscience & Biobehavioral Reviews, 111*, 84-94.

Gureckis, T. M., & Love, B. C. (2009). Short-term gains, long-term pains: How cues about state aid learning in dynamic environments. *Cognition*, *113*(3), 293-313.

Robinson, T. E., & Flagel, S. B. (2009). Dissociating the predictive and incentive motivational properties of reward-related cues through the study of individual differences. *Biological psychiatry*, *65*(10), 869-873.

Sutton, R. S. (1984). *Temporal credit assignment in reinforcement learning*. Doctoral dissertation, Department of Computer and Information Science, University of Massachusetts Amherst.