Learning to expect change: Volatility during early experience alters reward expectations in a model of interval timing

Nora C. Harhen (nharhen@uci.edu)

Aaron M. Bornstein (aaron.bornstein@uci.edu)
Department of Cognitive Sciences, University of California, Irvine, Irvine, CA 92697 USA

Abstract

An unpredictable early life environment can have enduring effects on mental health outcomes in adulthood. Despite widespread evidence for this relationship, it remains unclear what core mechanism links the two. Here we propose that early life unpredictability (ELU) shapes the development of temporal sequence representations. Critically, we show that this in turn produces impairments in reward sensitivity and learning, phenotypes that have been associated with anhedonia, a transdiagnostic symptom often observed in individuals with ELU. We formalize this hypothesis using a principled model of interval timing whose representations adjust with experience to support adaptive temporal predictions. The core observation is that initial unpredictability in timing produces broader, more imprecise temporal expectations. As a result, reward anticipation and learning are diminished. When we introduced agents with broader expectations into a stable environment, they showed a greater response to the omission of reward relative to its presence. This bias accords with negative attentional and mnemonic biases associated with anhedonia. In sum, we show that a single mechanism can explain a range of behaviors associated with anhedonia, offering insights into the role of temporal representations in reward learning and in the emergence of phenotypes linked to psychiatric disorders.

Keywords: early life unpredictability; reinforcement learning; interval timing; temporal representation

Introduction

Across development, brain circuits adapt to meet the demands of the environment. Concretely, sensory receptive fields are tuned to reflect the statistics of the early life environment, determining perceptual discrimination abilities in adulthood. Consistency is crucial to this maturation process. For functional circuits to form, the input statistics must be consistent (Li, Fitzpatrick, & White, 2006). It has recently been proposed that similar processes may occur in reinforcement and memory systems critically involved in associative learning (Birnie et al., 2020). This implies that the consistency or predictability of associations encountered early in life may shape the acquisition of associations later on.

Interactions with caregivers are one contributor to the associative statistics an infant encounters. For example, the infant behaves in some way and, normatively, the caregiver produces a consistent response to this behavior such that the infant can anticipate the response in the future. The timing between behavior and response is encoded and can be represented using a set of temporal receptive fields (TRFs) similar to receptive fields found in sensory areas. Instead of being tuned to visual angle or auditory pitch, these TRFs are sensitive to the time between associated stimuli and its consistency.

Caregivers vary in the valence and predictability of their responses. Most prior work has focused on the effect of valence on later child mental health outcomes. However, recent work has begun to examine how early life unpredictability, or ELU, might also contribute (Baram et al., 2012). Caregiver signals, if unpredictable, can result in anhedonia-like behaviors such as reduced experience of pleasure and motivation (Bolton et al., 2018). Importantly, anhedonia is a transdiagnostic symptom associated with several psychiatric disorders previously shown to be related to ELU (Glynn et al., 2019).

In the current work, we propose that TRFs are tuned to the unpredictability of timing in the environment, and these adaptations produce an anhedonic phenotype. We extend a principled computational model of interval timing (Ludvig, Sutton, & Kehoe, 2008) to examine how enhanced volatility during an early period of plasticity can, with minimal assumptions, affect later predictions of reward during maturity, when adaptation no longer occurs. With this model, we formally demonstrate that early unpredictability in timing and adaptation of temporal receptive fields to this timing can lead to an array of anhedonia-like symptoms. This includes an asymmetric response to reinforcement and omission despite no differences in the overall amount of reinforcement. This reproduces empirical findings that poor mental health outcomes can emerge from unpredictability in early life experience beyond what would be predicted from the overall number of adverse events (Glynn et al., 2019).

Methods

The Temporal-Difference model

Temporal-Difference (TD) models aim to accurately estimate the value of states in the world, V, in terms of the future rewards they predict. Time is explicitly represented in these models with a separate V for each time step, t, in a trial.

$$V^* = E[\sum_{k=1}^{\infty} \gamma^{k-1} r_{t+k}] \quad (1)$$

where r_t is the reward received at the current time step, and γ controls how heavily future rewards are discounted. Future rewards are less influential on V when γ is low. A TD agent

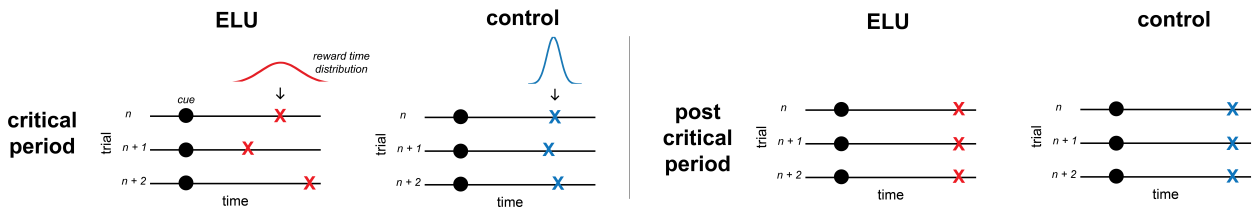

Figure 1: Two groups of agents, early life unpredictability (ELU) and control, learned to associate a cue and reward across two environments. The cue was partially reinforced in both environments — 75% of the time in the first and 50% in the second. During the first phase, both groups adapted their temporal receptive fields to the statistics of reward timing. The timing of reward delivery varied from trial to trial, differently for each group: The ELU group's timing was sampled from a much wider distribution relative to the controls. However, during the second phase, both groups received rewards at the exact same time on every reinforced trials.

learns V by an error driven learning rule. The estimate of V at the next time step is updated using the difference, δ_t, between the reward that was predicted (V_{t-1}) and what was actually received ($r_t + \gamma V_t$).

$$\delta_t = r_t + \gamma V_t - V_{t-1} \tag{2}$$

The Microstimulus model

All TD models explicitly represent time, but do so in various ways. Basic TD models use a complete-serial-compound (CSC) representation in which each time step is treated as independent from one another. The agent is assumed to have perfect knowledge of the time between cue and reward. This representation prohibits temporal generalization, creating issues in environments where the time between cue and reward varies. The *microstimulus* representation addresses this problem by relaxing its temporal markers (Ludvig et al., 2008). CSC's discrete markers are replaced with less precise microstimuli that allow for uncertainty to be represented. A stimulus, whether it be a neutral, rewarding, or aversive is assumed to leave behind a memory trace that decays with time. The trace is represented by a basis set of overlapping *temporal receptive fields* — Gaussian distributions whose standard deviations increase with the time after onset of the initial stimulus.

$$f(y,\mu,\sigma) = \frac{1}{\sqrt{2\pi}} e^{(-\frac{(y-\mu)^2}{2\sigma^2})} \tag{3}$$

A time step's value, V_t, is estimated as the weighted average of the microstimuli.

$$V_t = w_t^T x_t = \sum_{i=1}^{n} w_t(i) x_t(i) \tag{4}$$

This value is compared to the reward received. The error term, δ_t, adjusts the weights on the microstimuli, consequently updating the predicted value at the next time step.

$$w_{t+1} = w_t + \alpha \delta_t e_t \tag{5}$$

α is the learning rate controlling the time window over which trial to trial experiences are integrated. e_t is a vector containing each stimulus's eligibility traces.

$$e_t = \gamma \lambda e_t + x_t \tag{6}$$

Following the stimulus, its eligibility trace decays at a rate determined by γ and λ. γ is a discounting factor as above while λ controls the time window over which a stimulus can induce learning within a trial. For all simulations, we use the parameter settings from Ludvig et al, 2008 — $\alpha = 0.01$, $\gamma = 0.98$, $\lambda = 0.95$, $n = 50$, and $\sigma = 0.08$.

Simulating development

To model developmental changes in learning, we limit the period over which microstimuli weights can adapt to experience. We treat this as a critical period during which the temporal receptive fields are tuned to support accurate estimation of V. This adaptation process is designed to mimic the observed tuning of sensory receptive fields during analogous sensitive periods of development (Simoncelli & Olshausen, 2001).

We simulated two groups of agents learning cue-reward pairings across two phases (Figure 1). One group of agents, the early life unpredictability or ELU group, experienced a volatile environment in the first phase. Specifically, the delay between cue and reward considerably varied from trial to trial. The other group of agents, the control group, experienced relatively much less variation.

On each of the 1000 simulated trials, a cue was always presented at 100 milliseconds and there was a 75% probability of a reward following it. If a cue was reinforced on a trial, the timing of reward delivery was sampled from a normal distribution with μ set to 300 milliseconds for all agents while σ varied. For the ELU group, σ was sampled from a zero-truncated normal distribution with $\mu_{hyper,elu} = 10$ and $\sigma_{hyper,elu} = 3$. The control group experienced much less temporal variability with σ being sampled from a zero-truncated normal distribution with $\mu_{hyper,control} = 1$, $\sigma_{hyper,control} = 2$.

In the second phase, the weights could no longer adapt to the new environment. Thus, this phase was post the critical period. Both groups encountered another 1000 trials of learning to pair the same cue to a reward. On each trial, there was now a 50% probability of reward being presented following

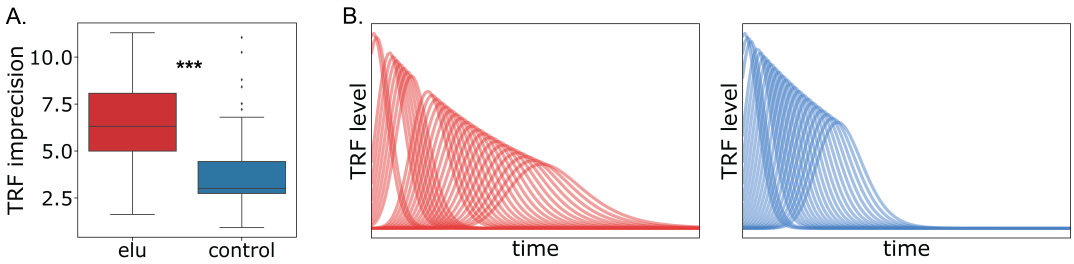

Figure 2: **A.** Temporal receptive field (TRF) imprecision was computed by taking a weighted average of the standard deviations of the temporal receptive fields following the critical period phase. The ELU group showed greater average temporal receptive field imprecision, a consequence of their more volatile experience during the critical period. **B.** Both groups' positively weighted temporal receptive fields. Recapitulating the results shown in panel A, the ELU group relied on more broadly tuned, less precise temporal receptive fields relative to the control group.

the cue. As before, the cue arrived at 100 ms. Reward timing was more stable in this environment with reward always arriving at 500 ms.

Here we focus on anhedonia, variously defined as the inability to experience and/or anticipate pleasure, as a symptom associated with many disorders observed to result following ELU. Following previous work, we model anhedonia as a reduced sensitivity to rewards and an impaired ability to learn from reinforcement (Huys, Pizzagalli, Bogdan, & Dayan, 2013). We asked if the simulated agents could exhibit these features of anhedonia from variability in reward timing alone, despite outcome valence being equated across groups.

Results

Critical Period

First, we examined how the initial environment shaped the tuning of temporal receptive fields by comparing the groups' microstimuli weights following the critical period. For each agent, we computed a temporal precision measure by taking a weighted average of the microstimuli's standard deviations. We found that the ELU group relied on more broadly-tuned receptive fields relative to the controls (Figure 2; t(198)$= -7.83$, $p < .0001$).

Prior work has demonstrated that early life unpredictability impedes learning from reinforcement (Birn, Roeber, & Pollak, 2017; Dillon et al., 2009). Thus, we examined whether the model could capture this. As a proxy for learning, we use prediction error magnitude. The more an agent has learned to associate a cue and a reward, the smaller their prediction error will be when a cue is reinforced with reward and the greater their prediction error will be when reward is omitted. To compare prediction errors between groups, we computed the median prediction error extremum for each agent. On reinforced trials, the maximum prediction error magnitude following the cue was taken while on omission trials, we took the minimum. We found that the ELU group demonstrated more extreme prediction errors relative to controls on reinforced trials (Figure 3, 4, t(198) $= 15.15$, $p < .0001$) but less extreme on omission trials (t(198) $= 6.09$, $p < .0001$). These results are consistent with the ELU group showing weaker learning under reinforcement. Critically, this is despite experiencing the same amount of reward on average as the control group (t(198)$= 0.67$, $p = 0.51$). This suggests that impaired reward learning, as observed in anhedonia, can emerge from experienced temporal volatility alone during a period of plasticity.

Early life unpredictability has also been shown to impair motivation (Hanson, Williams, Bangasser, & Peña, 2021). This may stem from a reduced expectation of reward. Thus, we compared the groups' expectation of value across time following the cue. The ELU group's value signal peaked early following the cue (mean = 156 ms; sd = 27) and slowly decayed, not reaching its minimum for several 100s of milliseconds following the cue (mean = 500 ms; sd = 1.4). This suggests if the reward is not received immediately, ELU individuals gradually grow less confident it will come at all. Conversely, the control group's signal peaked much later (mean = 265; sd = 19; t(198) $= 32.66$, $p < .0001$) but reached its minimum much sooner near the average reward time (Figure 5 mean = 373; sd = 89; t(198)$= -15.45$, $p < .0001$). In other words, control individuals increasingly anticipate the reward as its expected arrival time approaches.

Post Critical Period

During the second phase, the reward timing was consistent for both groups and the weights were no longer allowed to adapt. Under these conditions, the ELU group showed less extreme positive prediction errors relative to controls (Figure 6, t(198) $= -14.57$, $p < .0001$) but more extreme negative prediction errors (t(198) $= -8.13$, $p < .0001$), the opposite pattern as observed during the critical period.

To ensure our simulated agents' bias did not emerge from aggregating over the data, we computed an asymmetry index for each agent:

$$index = \frac{PE_+ - PE_-}{PE_+ + PE_-} \tag{7}$$

We found that the the ELU group had asymmetry indices that were in aggregate negative (t(199) $= -2.87$, $p = .005$)

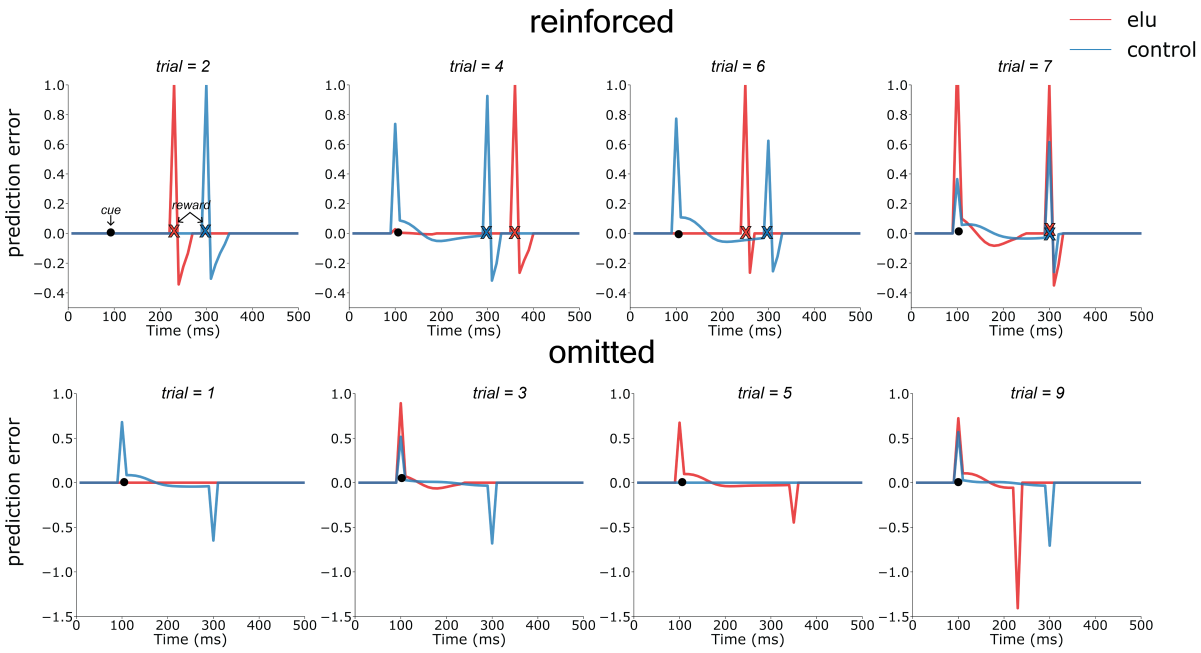

Figure 3: Critical period results - prediction error. Prediction error, δ, across time on trials where the cue was reinforced versus when it was omitted. For the ELU group, the timing of the large prediction error following the cue varies from trial to trial as a result of the reward timing varying. In contrast, the control group consistently experience a large prediction error near 300 ms.

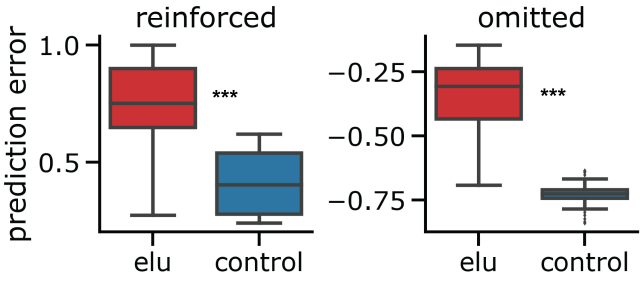

Figure 4: Critical period results - median prediction error extremum. For each trial, the extreme points of the prediction error was taken following the cue. For each agent, the measure was computed by taking the median over the trials' extremums. The ELU group showed larger predictions errors on trials where the cue was reinforced but weaker prediction errors when reward was omitted following the cue. Stars indicate significance of the test reported in the main text as follows: * $p < .05$, ** $p < .01$, *** $p < .001$.

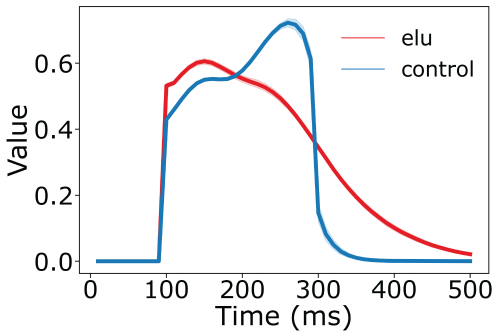

Figure 5: Critical period results - value. V, at each time step averaged across trials. The ELU group's value decreased following the cue while the control group's increased. Once the typical reward time was reached, the ELUs' value signal continued to steadily drop while the controls' did so quickly.

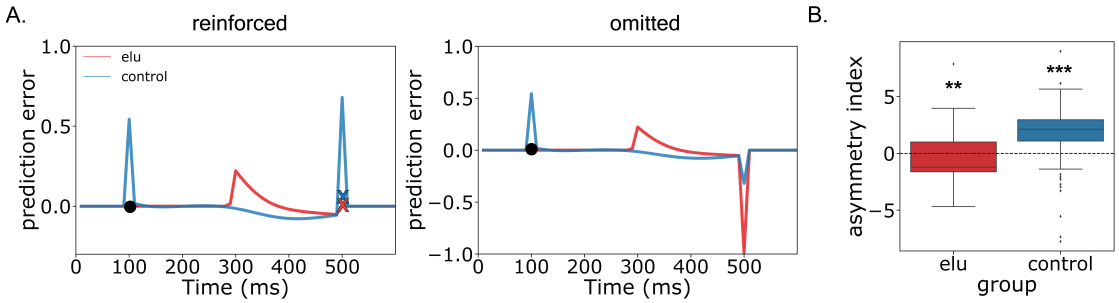

Figure 6: Post critical period results. **A.** Prediction errors at each time step for reinforced and omitted trials. On trials where the cue was reinforced, the control group showed larger and earlier cue-related and reward-related prediction errors relative to the ELU. On trials where reward was omitted, again, the control group showed larger and earlier cue-related prediction errors. However, for the reward-related prediction errors, the ELU groups' were larger. **B.**Asymmetry index. The ELU group displayed more extreme prediction errors on omission trials relative to reinforced while the control group showed the opposite pattern.

while the control group's were positive ($t(199) = 7.00$, $p < .0001$).

Discussion

Here, we've proposed a novel computational link between early life unpredictability and the emergence of anhedonia — the optimization of temporal representations to the early life environment. We assume that the volatility of the early life environment adaptively tunes temporal receptive fields in such a way that several behaviors associated with anhedonia — impaired learning from reinforcement reduced anticipation of reward, and a greater response to the omission of events — emerge.

These findings are consistent with behavioral outcomes observed in the laboratory and clinical settings. One representative such set of findings is of an asymmetric attentional bias in anhedonia. If we assume that attention increases with prediction error magnitude, then the ELU group were attentionally biased toward the omission outcome over the reinforced. Additionally, if we treat the omission of reward as a negatively valenced event and the presence of reward as positive, this suggests a negative attentional bias in the ELU group and positive bias in the controls, reproducing empirical findings (Dillon & Pizzagalli, 2018; Frank, 2004). Larger negative prediction errors may not only affect attention in the moment but also shape mood over the longer term (Eldar, Rutledge, Dolan, & Niv, 2016). Recurring negative prediction errors may give rise to the persistent negative mood that characterizes anhedonia (Dillon et al., 2009).

In the current work, we've interpreted the results while treating the outcome paired with the cue as a reward. However, the model is agnostic to whether the associated stimuli are neutral, rewarding, or aversive. Different outcome valences suggest different behavioral phenotypes. If the outcome is aversive, like a shock, rather than a reward, the ELU group's prolonged expectation of an outcome's appearance could produce a sort of "paranoia". The agent generalizes their expectation of the aversive event over a longer time period, producing a continual state of nervousness that aligns with symptoms of anxiety. If the outcome is neutral, impairments in reward learning become more general impairments in relational learning. This may explain memory deficits and alterations in hippocampal structure in ELU individuals (Granger et al., 2021; Molet et al., 2016) and its relationship with anhedonia. Prior work has suggested that anhedonia is characterized not only by the inability to experience pleasure in the moment but also the inability to recall past and anticipate future pleasurable experiences (Dillon & Pizzagalli, 2018).

Here we've only considered the mechanism under Pavlovian learning conditions. However, it suggests differences in ELU individuals' instrumental learning and action selection. The inability to accurately predict the timing of future outcomes diminishes an individual's perceived controllability of the environment, which has been implicated in psychiatric disorders such as anxiety (Bishop & Gagne, 2018).

Hidden-state inference models capture a similar idea as the microstimulus model at a different level of analysis (Starkweather, Babayan, Uchida, & Gershman, 2017). Often, the true state of the world is unknown or hidden and must be inferred from observations. This inference process is in part driven by prediction errors (Rouhani, Norman, Niv, & Bornstein, 2020), and by extension is more difficult in volatile environments. As a result, ELU individuals may infer fewer states in the world (or, analogously, more states in an environment where negative prediction errors predominate) and group their experiences accordingly as a result of this early volatility. We have previously shown that this assumption of reduced sensitivity with a hidden-state inference model can produce reduced exploration in a foraging task (N. C. Harhen & Bornstein, 2021), a behavior found in ELU populations (Lloyd, McKay, & Furl, 2022) , and may also explain why individuals who experience early life unpredictability are at higher risk of developing substance use disorders and relapsing following treatment (N. Harhen, Baram, Yassa, & Bornstein, 2021).

Our results highlight the key role time plays in shaping reinforcement learning and consequently its impact on behaviors associated with mental illness. The varied phenotypes that emerge from the same computations is consistent with the idea that the mechanism identified here has implications that extend beyond anhedonia. It suggests a common origin for a number of psychiatric disorders, potentially explaining their high co-morbidity rates. Further empirical research is needed to test the model's behavioral implications for early life unpredictability's impact on interval timing, and interval timing's relationship with psychiatric disorders.

Acknowledgements

This work was supported by a NARSAD Young Investigator Award from the Brain and Behavior Research Foundation to AMB. NCH was supported by a National Defense Science and Engineering Graduate fellowship. The authors thank Elliot Ludvig for providing the microstimulus model code, and Tallie Z Baram, Michael A Yassa, Steven J Granger, Steven V Mahler, and Sophia C Levis for helpful conversations.

References

Baram, T. Z., Davis, E. P., Obenaus, A., Sandman, C. A., Small, S. L., Solodkin, A., & Stern, H. (2012, September). Fragmentation and unpredictability of early-life experience in mental disorders. *Am. J. Psychiatry*, *169*(9), 907–915.

Birn, R. M., Roeber, B. J., & Pollak, S. D. (2017, December). Early childhood stress exposure, reward pathways, and adult decision making. *Proc. Natl. Acad. Sci. U. S. A.*, *114*(51), 13549–13554.

Birnie, M. T., Kooiker, C. L., Short, A. K., Bolton, J. L., Chen, Y., & Baram, T. Z. (2020, May). Plasticity of the reward circuitry after Early-Life adversity: Mechanisms and significance. *Biol. Psychiatry*, *87*(10), 875–884.

Bishop, S. J., & Gagne, C. (2018, July). Anxiety, depression, and decision making: A computational perspective. *Annu. Rev. Neurosci.*, *41*, 371–388.

Bolton, J. L., Molet, J., Regev, L., Chen, Y., Rismanchi, N., Haddad, E., ... Baram, T. Z. (2018, January). Anhedonia following Early-Life adversity involves aberrant interaction of reward and anxiety circuits and is reversed by partial silencing of amygdala Corticotropin-Releasing hormone gene. *Biol. Psychiatry*, *83*(2), 137–147.

Dillon, D. G., Holmes, A. J., Birk, J. L., Brooks, N., Lyons-Ruth, K., & Pizzagalli, D. A. (2009, August). Childhood adversity is associated with left basal ganglia dysfunction during reward anticipation in adulthood. *Biol. Psychiatry*, *66*(3), 206–213.

Dillon, D. G., & Pizzagalli, D. A. (2018, March). Mechanisms of memory disruption in depression. *Trends Neurosci.*, *41*(3), 137–149.

Eldar, E., Rutledge, R. B., Dolan, R. J., & Niv, Y. (2016, January). Mood as representation of momentum. *Trends Cogn. Sci.*, *20*(1), 15–24.

Frank, M. J. (2004). *Dynamic dopamine modulation of striato-cortical circuits in cognition: Converging neuropsychological, psychopharmacological and computational studies* Unpublished doctoral dissertation. Ann Arbor, United StatesUniversity of Colorado at Boulder.

Glynn, L. M., Stern, H. S., Howland, M. A., Risbrough, V. B., Baker, D. G., Nievergelt, C. M., ... Davis, E. P. (2019, April). Measuring novel antecedents of mental illness: the questionnaire of unpredictability in childhood. *Neuropsychopharmacology*, *44*(5), 876–882.

Granger, S. J., Glynn, L. M., Sandman, C. A., Small, S. L., Obenaus, A., Keator, D. B., ... Davis, E. P. (2021, February). Aberrant maturation of the uncinate fasciculus follows exposure to unpredictable patterns of maternal signals. *J. Neurosci.*, *41*(6), 1242–1250.

Hanson, J. L., Williams, A. V., Bangasser, D. A., & Peña, C. J. (2021). Impact of early life stress on reward circuit function and regulation. *Front. Psychiatry*, *12*, 1799.

Harhen, N., Baram, T. Z., Yassa, M. A., & Bornstein, A. M. (2021, May). Formalizing the relationship between early life adversity and addiction vulnerability: The role of memory sampling. *Biol. Psychiatry*, *89*(9), S189.

Harhen, N. C., & Bornstein, A. M. (2021). Structure learning as a mechanism of overharvesting. *Proceedings of the 19th International Conference on Cognitive Modeling*.

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013, June). Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biol. Mood Anxiety Disord.*, *3*(1), 12.

Li, Y., Fitzpatrick, D., & White, L. E. (2006, May). The development of direction selectivity in ferret visual cortex requires early visual experience. *Nat. Neurosci.*, *9*(5), 676–681.

Lloyd, A., McKay, R. T., & Furl, N. (2022, January). Individuals with adverse childhood experiences explore less and underweight reward feedback. *Proc. Natl. Acad. Sci. U. S. A.*, *119*(4).

Ludvig, E. A., Sutton, R. S., & Kehoe, E. J. (2008, December). Stimulus representation and the timing of reward-prediction errors in models of the dopamine system. *Neural Comput.*, *20*(12), 3034–3054.

Molet, J., Maras, P. M., Kinney-Lang, E., Harris, N. G., Rashid, F., Ivy, A. S., ... Baram, T. Z. (2016, December). MRI uncovers disrupted hippocampal microstructure that underlies memory impairments after early-life adversity. *Hippocampus*, *26*(12), 1618–1632.

Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2020, October). Reward prediction errors create event boundaries in memory. *Cognition*, *203*, 104269.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.*, *24*, 1193–1216.

Starkweather, C. K., Babayan, B. M., Uchida, N., & Gershman, S. J. (2017, April). Dopamine reward prediction errors reflect hidden-state inference across time. *Nat. Neurosci.*, *20*(4), 581–589.