Applying Reinforcement Learning to Rodent Stress Research

Clara Liao[1] and Alex C. Kwan[1,2,3] (iD)

Abstract

Rodent models are an invaluable tool for studying the pathophysiological mechanisms underlying stress and depressive disorders. However, the widely used behavioral assays to measure depressive-like states in rodents have serious limitations. In this commentary, we suggest that learning tasks, particularly those that can be analyzed with the framework of reinforcement learning, are ideal for assaying reward processing deficits relevant to depression. The key advantages of these tasks are their repeatable, quantifiable nature and the link to clinical studies. By optimizing the behavioral readout of stress-induced phenotypes in rodents, a reinforcement learning-based approach may help bridge the translational gap and advance antidepressant discovery.

Introduction

Depression is a debilitating and prevalent disorder characterized by a constellation of symptoms including low mood, amotivation, and cognitive impairments. Anhedonia manifests in depressed subjects due to blunted responsiveness to positive outcomes, suggesting underlying neural dysregulations in reward-guided decision making.[1,2] Rodents can be an invaluable tool for modeling this particular aspect of the disorder, allowing researchers to manipulate the brain at molecular, genetic, and circuit levels to gain insight into the pathophysiology. By studying stress paradigms as rodent models for depression, there is hope for determining the neural mechanisms underlying maladaptive behaviors and identifying novel antidepressants.

However, efforts to translate findings from animals to humans have been hampered by limitations of current rodent behavioral assays. Traditionally, the depressive- and anxiety-like states in rodents are evaluated by administering a battery of tests including tail suspension, forced swim, sucrose preference, urine sniffing, and others. Although these tests have undoubtedly provided important insights into the etiology of stress-induced dysfunctions, the behavioral assays have several notable shortcomings. One, many assays cannot be repeatedly administered because animals can develop coping strategies against simple challenges. Two, measurements are prone to subjective error because behavioral responses are often scored based on visual inspection. Three, because the assays have no clinical counterparts in humans, any alterations – measured as immobility duration, sucrose consumed, time spent sniffing, and so forth – have to be interpreted anthropomorphically to relate to depressive-like states in humans.[3] Due to these shortcomings, such behavioral assays should not be used as the sole readout in experiments, as they are susceptible to detecting false positives. Thus, there is a need to expand the battery of tests for evaluating rodent models for depression, specifically including behavioral assays that are more repeatable, more quantitative, and more relatable to human behaviors.

[1]Interdepartmental Neuroscience Program, Yale University School of Medicine, New Haven, CT, USA
[2]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA
[3]Department of Neuroscience, Yale University School of Medicine, New Haven, CT, USA

Corresponding Author:
Alex C. Kwan, Yale University School of Medicine, New Haven, CT 06511, USA.
Email: alex.kwan@yale.edu

In this commentary, we suggest that reward-based learning tasks – particularly those that can be analyzed within the framework of reinforcement learning – are ideal for characterizing reward processing dysfunctions in rodent models for depressive-like behaviors.

What is reinforcement learning?

It is a natural and adaptive process for animals and humans to select actions that will maximize rewarding outcomes. This requires the subject to learn from past actions: choices that result in a positive outcome should be repeated, whereas choices that yield lower rewards than expected, or even punishment, should be avoided. It is important to note that, if the environment is stable (actions always lead to the same outcomes), a subject can quickly grasp the best options and no longer has an incentive to learn. By contrast and more in line with real-life situations, if the environment is *dynamic* (action-outcome contingencies can change over time) and *uncertain* (the same action can lead probabilistically to different outcomes), a subject must continually learn from prior experiences and outcomes to adapt to the changing environment. Rats and mice are adept at such dynamic adjustments in foraging tasks.[4–7]

Reinforcement learning is a computational framework for understanding the learning that occurs in a dynamic and uncertain environment. It provides a set of equations that fully describe how a subject would perform in a reward-based learning task. The equations are fitted to empirically measured behavioral data, and parameters of the equations are extracted. Subsequently these learning parameters and equations can be applied to predict how the subject would perform in other tasks and learning situations. Furthermore, distinct learning strategies can be encapsulated by posing different sets of equations (e.g., Q-learning, Bayesian updating, etc.). The fits to empirical data can be compared rigorously through model selection to determine the learning strategy that is most likely employed by the subject.

More repeatable

An advantage of reward-based learning tasks is that the assays are is repeatable. By design, each session involves upwards of several hundreds of trials. Animals can be tested repeatedly across multiple sessions, because the environment is dynamic and subjects have to continually adapt throughout the assay. This is in stark contrast to traditional behavioral tests where repeated measurements often lead to variable outcomes because animals can develop coping strategies in a stable task. For example, in forced swim tests, a shift from escape behavior to immobility is interpreted as a readout of behavioral despair. However, over successive tests, animals can learn to cope in the task by floating – an alternative, confounding strategy that appears as sustained immobility, leading to inaccurate results over repeated measurements.[8]

The ability to take repeated measurements in a behavioral task is advantageous because it allows for within-subject design, which has greater statistical power than between-subject design. Moreover, the same animals can be assessed before, during, and after stress exposures or pharmacological manipulations, enabling researchers to identify the latency and duration of stress and drug effects and study the associated neuronal changes throughout the time course. Studies that have investigated the longitudinal effects of chronic stress have revealed that successive stress episodes are associated with accumulating deficits in reward-guided actions, which are accompanied by progressive modifications in neuronal activity.[9–11]

More quantitative

A single session of a reward-based learning task typically consists of several hundreds of trials. The large data set ensures accuracy in the fitting of the reinforcement learning equations and confidence in the extraction of defined learning parameters. The learning parameters have predictive power; providing a quantitative value assigned to the impact of the experimental manipulation that can be compared across studies (Figure 1). Take the example of a pharmacological manipulation of Drug A that causes the learning rate to drop by 10%: the alteration in the subject's performance in any reward-based learning task can be simulated computationally. Next, if another compound, Drug B, reduces learning rate by 20%, then its impact can also be simulated and the difference in their efficacies in altering a subject's decision tendency can be determined exactly. In other words, researchers can make quantitative statements about changes due to experimental manipulations. By contrast, with traditional behavioral tests, the metrics are not easily comparable. For example, if Drugs A and B reduce immobility in tail suspension by 10 versus 20%, or in tail suspension by 10% versus in forced swim test by 20%, what does that say about the drugs' relative efficacy? It would be unclear if that should be interpreted as a small or big difference. Thus, the quantitative parameterization of behaviors afforded by reinforcement learning is a principled way to assess stress-induced alterations.

More relatable to human behaviors

Decision-making with uncertainty and in a dynamic environment not only requires continual learning in animals, but is also a non-trivial problem for humans.
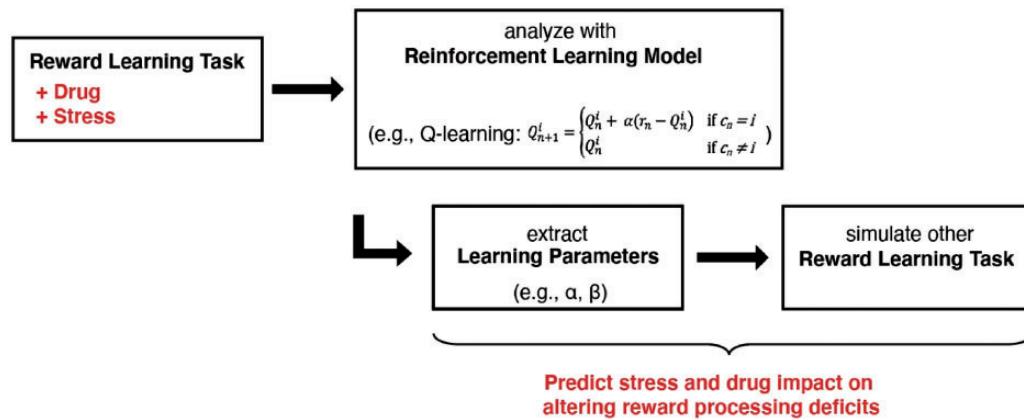
Reward Learning Task
+ Drug
+ Stress

analyze with
Reinforcement Learning Model

$$(e.g.,\ Q\text{-learning}:\ Q_{n+1}^i = \begin{cases} Q_n^i + \alpha(r_n - Q_n^i) & \text{if } c_n = i \\ Q_n^i & \text{if } c_n \neq i \end{cases})$$

extract
Learning Parameters
$(e.g.,\ \alpha,\ \beta)$

simulate other
Reward Learning Task

**Predict stress and drug impact on
altering reward processing deficits**

Figure 1. Schematic depicting how applying reinforcement learning can provide quantitative parameterization of stress or drug-induced alterations of behavior.

Therefore, humans can be tested on similar reward-based learning tasks and reinforcement learning can be likewise applied to analyze the decision-making process. This approach has been used to study humans under stress or suffering from depression symptoms. For example, stressed subjects favor habitual behaviors at the expense of goal-directed actions in instrumental learning.[12] Patients with major depressive disorder show blunted responses to feedback information, including a hyposensitivity to reward and deficits in response to negative feedback.[2,13] Based on these results and other work, it has been argued on theoretical grounds that defects in specific learning parameters in reinforcement learning may capture aspects of depression.[14,15] Indeed, recent empirical studies found that depression and use of antidepressants are associated with altered learning parameters.[16,17] Looking forward, the findings in humans can be studied in greater detail in rodent models, where researchers have the ability to investigate which experimental manipulations or interventions can influence those learning parameters, ultimately providing insight into the pathophysiology of the disorder. Therefore, reinforcement learning presents a potential translational link between rodent stress models and clinical studies.

Limitations and outlook

It is important to note the challenges ahead for using reinforcement learning to study rodent models for depression. First of all, it has not yet been fully established how chronic stress affects rodents' decisions in tasks involving an uncertain and dynamic environment. However, several lines of evidence suggest a reasonable anticipation of deficits. For example, rats subjected to chronic unpredictable stress become insensitive to changes in outcome value in operant devaluation tests.[18] With social stress, defeated animals had diminished flexibility, being unable to shift their behavior in response to switches in action-outcome contingency.[11] Furthermore, conventional and fast-acting antidepressants induce notable effects on performance of wild type rats in a probabilistic reward learning task.[19]

It is crucial to recognize that reward processing deficits represent only one dimension of human depression. Given the heterogeneity and range of impairments associated with depression, modeling the disorder in rodents is challenging and unlikely to recapitulate the full extent of symptoms in humans.[3] With the limitations in mind, there remain definitive advantages for pursuing reward-learning tasks based on the framework of reinforcement learning as novel assays to evaluate rodent models of depression. The more repeatable, quantitative, and relatable approach promises to facilitate the translation of findings in animal models to improve diagnostics and identify new treatment options for depression.

ORCID iD

Alex C. Kwan https://orcid.org/0000-0003-2169-1667

References

1. Eshel N, Roiser JP. Reward and punishment processing in depression. *Biol Psychiatry*. 2010;68:118–124.

2. Pizzagalli DA, Iosifescu D, Hallett LA, Ratner KG, Fava M. Reduced hedonic capacity in major depressive disorder: evidence from a probabilistic reward task. *J Psychiatr Res*. 2008;43:76–87.

3. Nestler EJ, Hyman SE. Animal models of neuropsychiatric disorders. *Nat Neurosci*. 2010;13:1161–1169.

4. Ito M, Doya K. Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *J Neurosci*. 2009;29:9861–9874.

5. Groman SM, Keistler C, Keip AJ, Rajagopalan AE, Cressy JI, Cohen JY. Orbitofrontal circuits control multiple reinforcement-learning processes [published online ahead of print June 30, 2019]. *Neuron*. doi:10.1016/j.neuron.2019.05.042

6. Bari BA, Grossman CD, Lubin EE, Rajagopalan AE, Cressy JI, Cohen JY. Stable representations of decision variables for flexible behavior processes [published online ahead of print July 10, 2019]. *Neuron*. doi:10.1016/j.neuron.2019.06.001

7. Hattori R, Danskin B, Babic Z, Mlynaryk N, Komiyama T. Area-specificity and plasticity of history-dependent value coding during learning. *Cell*. 2019;177:1858.e1815–1872.e1815.

8. Bogdanova OV, Kanekar S, D'Anci KE, Renshaw PF. Factors influencing behavior in the forced swim test. *Physiol Behav*. 2013;118:227–239.

9. Donahue RJ, Muschamp JW, Russo SJ, Nestler EJ, Carlezon WA Jr. Effects of striatal DeltaFosB overexpression and ketamine on social defeat stress-induced anhedonia in mice. *Biol Psychiatry*. 2014;76:550–558.

10. Der-Avakian A, Mazei-Robison MS, Kesby JP, Nestler EJ, Markou A. Enduring deficits in brain reward function after chronic social defeat in rats: susceptibility, resilience, and antidepressant response. *Biol Psychiatry*. 2014;76:542–549.

11. Barthas F, Hu MY, Siniscalchi MJ, et al. Cumulative effects of social stress on reward-guided actions and prefrontal cortical activity. *Biol Psychiatry*. 2020;88:541–553.

12. Schwabe L and Wolf OT. Stress prompts habit behavior in humans. *J Neurosci*. 2009;29:7191–7198.

13. Elliott R, Sahakian BJ, Herrod JJ, Robbins TW, Paykel ES. Abnormal response to negative feedback in unipolar depression: evidence for a diagnosis specific impairment. *J Neurol Neurosurg Psychiatry*. 1997;63:74–82.

14. Chen C, Takahashi T, Nakagawa S, Inoue T, Kusumi I. Reinforcement learning in depression: a review of computational research. *Neurosci Biobehav Rev*. 2015;55:247–267.

15. Huys QJ, Daw ND, Dayan P. Depression: a decision-theoretic analysis. *Annu Rev Neurosci*. 2015;38:1–23.

16. Mukherjee D, Lee S, Kazinka R, Satterthwaite TD, Kable JW. Multiple facets of value-based decision making in major depressive disorder. *Sci Rep*. 2020;10:3415.

17. Michely J, Eldar E, Erdman A, Martin IM, Dolan RJ. SSRIs modulate asymmetric learning from reward and punishment. *bioRxiv*. 2020. doi:10.1101/2020.05.21.108266

18. Dias-Ferreira E, Sousa JC, Melo I, et al. Chronic stress causes frontostriatal reorganization and affects decision-making. *Science*. 2009;325:621–625.

19. Wilkinson MP, Grogan JP, Mellor JR, Robinson ESJ. Comparison of conventional and rapid-acting antidepressants in a rodent probabilistic reversal learning task. *Brain Neurosci Adv*. 2020;4:2398212820907177.