Biological Psychiatry: CNNI

Commentary

A Computational Modeling Approach Supports Negative Reinforcement Theories of Addiction

Jennifer L. Stewart

An action (drug use) repeatedly followed by addition of a positive outcome (euphoria or high) or removal of a negative outcome (withdrawal symptoms or negative affect) will increase the likelihood of this action in the future. The former case is an example of positive reinforcement and the latter case is an example of negative reinforcement. Motivation to take this action (drug use) may depend on the evaluation of the user's current internal bodily state and how much it differs from the expected bodily state, otherwise known as prediction error (for example, "I expected to feel good but I don't" or "I'm starting to feel bad when I shouldn't" versus "I feel better than I thought I would") (1,2). Prediction errors are considered positive when one is in a better state than expected and are considered negative when one is in a worse state than expected. Over the course of addiction, users shift from a pattern of drug consumption fueled by positive reinforcement or approach toward rewarding drug effects to one driven by negative reinforcement or avoidance of aversive mood, stress, and drug withdrawal states (3). This change in reinforcement mechanisms appears to parallel the remapping of striatal dopaminergic circuitry, wherein once-exaggerated responses to drug highs within this region of the brain are blunted in favor of heightened responses to drug habits (4). In other words, whereas individuals initially use a drug because it makes them feel good, over time they use the drug to avoid feeling bad, and the striatum is implicated in this switch as well as in the coding of prediction errors (3,5).

Few human studies have examined how brain and behavior mechanisms implicated in reinforcement processes change as a function of current drug use state (such as intoxication, satiation, and withdrawal/deprivation) within, as well as across, individuals who are addicted to drugs. Understanding how the current state of the individual (feeling good or feeling bad) impacts additional prediction error coding and resultant craving/urges to use drugs could help us develop more effective interventions for reducing relapse and enhancing long-term recovery from drug addiction. In addition, the inclusion of two statistical approaches may inform the conclusions we draw from future work in this area. Addiction samples often possess substantial heterogeneity in drug use chronicity, recency, and comorbidity. Within-subject study designs enable participants to serve as their own comparator for one drug state versus another, allowing for the crucial separation of variance attributable to individual differences versus variance attributable to drug state differences. Moreover, advances in computational psychiatry enable researchers to use both data- and theory-driven mathematical algorithms to further our understanding of complex relationships between the brain, behavior, and clinical symptoms implicated in reinforcement processes (6,7).

In the current issue of *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, Wang *et al.* (8) present findings from an innovative study that used a within-subjects design and a data-driven machine learning approach to identify state-dependent changes in reinforcement processes within the context of cocaine addiction. Men with current cocaine dependence (n = 22) completed a two-armed bandit task during functional magnetic resonance imaging under two counterbalanced conditions: 1) after using cocaine as usual (satiation) and 2) after 3 days of no cocaine use, verified by a urine screen (deprivation). On each trial of the task, participants chose between two actions designated by pictures. One action was more likely to result in high money loss, whereas the other action was more likely to result in low money loss over the course of a block of trials, although participants were not informed of these action-outcome associations ahead of time. Multiple blocks of trials were presented with new pictures for each block to determine how successful people were at learning to select the action leading to the least loss (the greatest reward). Addiction models posit that in a drug-deprived state, users will be motivated to take action (use drugs) to remove a negative outcome (withdrawal or feeling bad) (1–3), and therefore Wang *et al.* (8) hypothesized that cocaine deprivation would be linked to better avoidance of the high-loss action (which translates into greater positive prediction error, or performing better than expected) than cocaine satiation, and that this degree of avoidance would be positively associated with striatal activation.

Wang *et al.* (8) used a machine learning algorithm, Q-learning (9), to build a model of optimal action-outcome selection that was then applied to the two-arm bandit behavioral data in this sample. Q-learning evaluates the relationships between an agent (computer), an action (two stimulus options for the two-armed bandit), and the two possible outcomes (low loss and high loss), assuming that the goal of the agent is to maximize total future reward. The Q-learning algorithm assumes that the agent knows little about action-outcome associations at the start of the task and, as a function of each incremental trial, updates a Q-table, or model, indicating the maximum expected future reward that the agent will earn if a particular action is taken for a particular outcome. Given the Q-table results, a positive prediction error was then calculated for each participant for each trial, representing how much better the actual reward outcome was than expected. Moreover, a positive learning rate was calculated for each participant per trial, showing how much influence the positive prediction error

SEE CORRESPONDING ARTICLE ON PAGE 291

  https://doi.org/10.1016/j.bpsc.2019.01.003
 ISSN: 2451-9022

from the current trial had in updating the expected reward value of future trials. Negative prediction error and learning rate estimates were also calculated for each person, reflecting how much the actual reward outcome was worse than expected and how much this influenced future trials. These prediction error and learning rate metrics provide much more nuanced information about reward updating over time than just the percentage of correct trials achieved during the task.

These behavioral prediction errors and learning rates were correlated with functional magnetic resonance imaging signals in the striatum for each trial per participant per each condition (cocaine satiation vs. deprivation), and then the findings were compared between conditions. Consistent with hypotheses, cocaine deprivation produced 1) a higher positive learning rate than cocaine satiation and 2) a positive correlation between positive prediction error and striatal activation similar to a sample of healthy male comparison subjects (this correlation was absent for cocaine satiation). No findings emerged for negative prediction error or learning rate. Finally, and importantly, Wang *et al.* (8) demonstrated that heightened positive prediction errors in the striatum during cocaine deprivation mediated, or accounted for, the relationship between greater years of cocaine use and higher desire for cocaine during deprivation.

These findings support the idea that within the context of chronic drug use a drug withdrawal state is associated with a heightened avoidance of negative outcomes via striatal signaling as well as greater drug craving—a situation that could drive relapse. This study crucially implicates negative reinforcement mechanisms in drug addiction and points to potential modulation of these mechanisms during detoxification/withdrawal states to potentially enhance the chance of prolonged abstinence. This study sample was small and was composed solely of men, and therefore additional studies are warranted to replicate and extend these results to women as well as individuals addicted to other classes of drugs, such as opioids, alcohol, and marijuana, thereby enhancing generalizability.

Acknowledgments and Disclosures

This work was supported by The William K. Warren Foundation.

The author reports no biomedical financial interests or potential conflicts of interest.

Article Information

From the Laureate Institute for Brain Research, Tulsa, and the Department of Community Medicine, University of Tulsa, Tulsa, Oklahoma.

Address correspondence to Jennifer L. Stewart, Ph.D., Laureate Institute for Brain Research, 6655 Yale Ave, Tulsa, OK 74136; E-mail: jstewart@laureateinstitute.org.

Received Jan 16, 2019; accepted Jan 16, 2019.

References

1. Paulus MP, Tapert SF, Schulteis G (2009): The role of interoception and alliesthesia in addiction. Pharmacol Biochem Behav 94:1–7.
2. Paulus MP, Stewart JL (2014): Interoception and drug addiction. Neuropharmacology 76:342–350.
3. Koob GF, Volkow ND (2016): Neurobiology of addiction: A neurocircuitry analysis. Lancet Psychiatry 3:760–773.
4. Berridge KC, Robinson TE (2016): Liking, wanting, and the incentive-sensitization theory of addiction. Am Psychol 71:670–679.
5. McClure SM, Berns GS, Montague PR (2003): Temporal prediction errors in a passive learning task activate human striatum. Neuron 38:339–346.
6. Huys QJ, Maia TV, Frank MJ (2016): Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci 19:404–413.
7. Huys QJM, Maia TV, Paulus MP (2016): Computational psychiatry: From mechanistic insights to the development of new treatments. Biol Psychiatry Cogn Neurosci Neuroimaging 1:382–385.
8. Wang JM, Zhu L, Brown VM, De La Garza II R, Newton T, King-Casas B, Chiu PH (2019): In cocaine dependence, neural prediction errors during loss avoidance are increased with cocaine deprivation and predict drug use. Biol Psychiatry Cogn Neurosci Neuroimaging 4:291–299.
9. Watkins CJCH, Dayan P (1992): Q-learning. Machine Learning 8:279–292.