Review

Biological Psychiatry: CNNI

Model-Free Temporal-Difference Learning and Dopamine in Alcohol Dependence: Examining Concepts From Theory and Animals in Human Imaging

Quentin J.M. Huys, Lorenz Deserno, Klaus Obermayer, Florian Schlagenhauf, and Andreas Heinz

ABSTRACT

Dopamine potentially unites two important roles: one in addiction, being involved in most substances of abuse including alcohol, and a second one in a specific type of learning, namely model-free temporal-difference reinforcement learning. Theories of addiction have long suggested that drugs of abuse may usurp dopamine's role in learning. Here, we briefly review the preclinical literature to motivate specific hypotheses about model-free temporal-difference learning and then review the imaging evidence in the drug of abuse with the most substantial societal consequences: alcohol. Despite the breadth of the literature, only a few studies have examined the predictions directly, and these provide at best inconclusive evidence for the involvement of temporal-difference learning alterations in alcohol dependence. We discuss the difficulties of testing the theory in humans, make specific suggestions, and close with a focus on the interaction with other learning mechanisms.

Keywords: Alcohol, Computational psychiatry, Cue reactivity, Dopamine, Habits, Model-free, Reinforcement learning, Ventral striatum

http://dx.doi.org/10.1016/j.bpsc.2016.06.005

There is substantial evidence pointing to a role for dopamine in both addiction and learning, which naturally raises the question of whether dopamine's role in addiction is mediated via its role in learning. Most addictive substances result in dopaminergic release in the ventral striatum (VS) (1–5), where dopamine signals are at the center of biologically increasingly detailed (6) model-free temporal difference (MFTD) accounts of how the brain instantiates iterative learning from reinforcements (7). Given evidence that phasic dopamine signals have a causal role in learning (8), it is reasonable to expect that addictive substances might exert their nefarious effect in part by subverting dopamine's role in MFTD learning (9,10).

It is unclear to what extent such a theory is supported by existing evidence. Here, we therefore examine findings in one of the societally most important drugs of abuse (11,12) with a huge treatment gap (13): alcohol. We start with a theoretical overview, mapping the valuation of stimuli onto incentive salience and sign-tracking theories, and the valuation of actions onto habitization theories. We then review the relevant imaging literature in humans and close with a discussion of outstanding issues and the limitations of existing tests of MFTD theories in humans.

MODEL-FREE TEMPORAL DIFFERENCE LEARNING

Stimulus Values and Incentive Salience

One influential account of addiction builds on the finding that stimuli paired with dopamine release or stimulation acquire incentive salience (14,15), becoming 1) desirable, 2) reinforcing in their own right, and 3) motivating. These are typical of drug-associated stimuli and might thus contribute to both development and maintenance of addictive states. Anecdotally, patients often report relapsing after encountering alcohol-related stimuli.

Model-free prediction-error learning (16) iteratively updates reward expectation values V with a prediction error that measures the discrepancy with the actually obtained reward r:

$$V \leftarrow V + \alpha(r - V).$$

The value V is a running average of experienced reinforcements that summarizes past reinforcement experience. In MFTD learning, the prediction error incorporates changes in expected rewards induced by reward-predicting stimuli (16). When cues predictive of reward occur unexpectedly, a prediction error proportional to this expectation is elicited.

MFTD valuation of stimuli s results in Pavlovian stimulus values V(s) that capture the three core aspects of incentive salience (17,18). They become desirable in that approaching stimuli with positive value V(s) is formally optimal (16). Because the specifics of how and when reinforcement happened are discarded, the desirability becomes separated from the details of past experience. Second, temporal difference values capture secondary reinforcement because a positive change in reward expectation can formally substitute for actual rewards.

ISSN: 2451-9022

Third, the motivating aspects (19) are captured by the fact that expectations of reward determine the optimal rate of action (20), although notably this is not specific to MFTD values. Finally, the delay in adapting values to reflect the current rewards provides one account for why wanting the drug (captured by *V*) is distinct from liking it (the immediate reward *r* experienced on consumption) and suggests one reason why wanting may persist beyond liking (21).

Stimulus Values and Sign-Tracking

There is substantial individual variation in Pavlovian conditioning. When a light in one corner of a box predicts food delivery in the other corner, sign-tracking rats will come to approach the light conditioned stimulus (CS) and wait there until delivery of the unconditioned stimulus (US), e.g., food. Goal-trackers, in contrast, move to the food delivery site immediately. Phasic dopamine levels in the VS behave like MFTD prediction errors in sign-trackers only, and only in them can learning the CS-US relation be blocked by dopaminergic antagonists (22,23). Hence, sign-trackers rely on dopaminergically mediated MFTD learning, whereas goal-trackers do not. Only in sign-trackers does the CS acquire incentive salience.

The link to addiction comes through animals selectively bred to show high or low responsivity to novelty (24). The animals selectively bred to show high responsivity to novelty are preferentially sign-trackers for natural rewards and show a broad range of addiction-like features (25). They respond more to cocaine acutely and show more locomotor sensitization effects (26), show stronger drug-taking acquisition (27), work harder for cocaine (28), seek cocaine when it is no longer available (29), are more impulsive on a range of measures (29), and have reduced dopamine D_2 receptor (D_2R) availability, also implicated in human addiction (30–32). Cocaine cues lead to escalation and reinstatement after extinction in sign-trackers but not in goal-trackers (33). Alcohol releases dopamine (34), and exposure to sign-tracking paradigms in adolescence increases sign-tracking and ethanol intake in adulthood (35). In humans, effects of Pavlovian stimuli on instrumental behavior have been confirmed, and in one study general Pavlovian-to-instrumental transfer effects predicted to relapse risk in alcohol use disorder (AUD) patients (36), although differences between goal- and sign-trackers have not yet been explored with respect to addiction.

Values, Habits, and Devaluation

Alcohol intake is suggested to have habitual components (37–39) because substance use persists despite obvious harmful consequences. Habits are defined through devaluation insensitivity (40), whereby behavior will continue even though the outcome of the action is no longer consumed if available freely. Habits contrast with goal-directed behavior, where the action will only be performed if the action's goal is desirable. MFTD values, be they about states or behaviors, capture devaluation insensitivity because they rest entirely on past experiences about how actions lead to outcomes, and they are not updated by information purely about the outcome itself (41) until the association between the state or action and the revalued outcome has been experienced.

Phasic dopaminergic signals are present during instrumental learning (42), and dopamine (43), the dorsolateral striatum, and the infralimbic cortex are required for habit formation both for natural rewards (44,45) and for drugs such as alcohol (46). Similar to habits, sign-tracking itself is resistant to devaluation of the outcome, whereas goal-tracking is not (47), and over extended training goal-tracking for alcohol gives way to sign-tracking (48), suggesting a similarity to a MFTD valuation process.

Exposure to stimulants or alcohol speeds up habit formation [for drug or natural rewards (46,49)]. Furthermore, D_2R antagonism, putatively modeling the reduction in D_2R availability (30–32), further promotes this (50), and dopaminergic signals shift from ventral to dorsal striatum with progression of the habitization (51).

IMAGING MFTD PROCESSES IN ALCOHOLISM

Several features must be satisfied to establish the presence of MFTD learning signals (7,52). Unpredicted rewards and unpredicted changes in reward expectation should result in a positive signal proportional to the difference between reward and expectation or expectation change. Responses to predictable rewards should decrease over repetitions, whereas responses to neutral stimuli predicting rewards reliably should increase over the course of learning. Unexpected omission of an expected reward should result in a negative signal. MFTD signals should not be sensitive to devaluation.

The responses should be visible in dopaminergic target region blood oxygen level–dependent (BOLD) measurements (53,54). For Pavlovian (cue) processes, this should be in the VS (22,55) that receives a strong dopaminergic projection (56). For habitual (action) processes, the signals may arise in the dorsal striatum (51,55,57).

There are two different ways in which MFTD processes might contribute to the development of dependence. Alcohol might specifically affect MFTD learning for stimuli or behaviors associated with alcohol (described in MFTD Processes and Alcohol Cues). Alternatively, a predisposition toward MFTD learning observable also in nondrug scenarios may predispose toward alcohol addiction (described in MFTD Processes and Nonalcoholic Rewards).

MFTD Processes and Alcohol Cues

Alcohol may specifically usurp MFTD processes to engender particularly powerful learning in situations associated with it. Definite evidence for this would require the learning process to be observed longitudinally over the course of the development of addiction. Cross-sectional examination of the end-result of learning, that is, responses to putative CSs, is weaker. Nevertheless, on the basis of the features of MFTD learning noted above, the following criteria should be met to support the involvement of MFTD processes:

- 1a: Responses to drug CSs should be more pronounced among individuals who have developed an addiction than among those who have not.
- 1b: Unexpected presentations of drug CSs should be accompanied by phasic dopaminergic release in the VS for Pavlovian settings and either in the ventral or dorsal striatum for instrumental settings (55).