HHS Public Access

Author manuscript

Curr Opin Behav Sci. Author manuscript; available in PMC 2021 August 19.

Published in final edited form as:

Curr Opin Behav Sci. 2021 April ; 38: 40–48. doi:10.1016/j.cobeha.2020.08.007.

Computational theory-driven studies of reinforcement learning and decision-making in addiction: What have we learned?

Maëlle C.M. Gueguen[1], **Emma M. Schweitzer**[1,2], **Anna B. Konova**[1,*]

[1]Department of Psychiatry, University Behavioral Health Care, & the Brain Health Institute, Rutgers University—New Brunswick, Piscataway, USA;

[2]Graduate Program in Cell Biology & Neuroscience, Rutgers University—New Brunswick, Piscataway, USA.

Abstract

Computational psychiatry provides a powerful new approach for linking the behavioral manifestations of addiction to their precise cognitive and neurobiological substrates. However, this emerging area of research is still limited in important ways. While research has identified features of reinforcement learning and decision-making in substance users that differ from health, less emphasis has been placed on capturing addiction cycles/states dynamically, within-person. In addition, the focus on few behavioral variables at a time has precluded more detailed consideration of related processes and heterogeneous clinical profiles. We propose that a longitudinal and multidimensional examination of value-based processes, a type of dynamic "computational fingerprint", will provide a more complete understanding of addiction as well as aid in developing better tailored and timed interventions.

Introduction

Reinforcement learning and decision-making—collectively, "value-based decision-making" [1]—are integral to adaptive behavior in everyday life. Value-based decision-making comprises a feedback loop whereby the values of candidate actions are learned and updated through experience, and used to guide behavior that maximizes utility (and minimizes disutility). Disruption in value-based decision-making is considered a key factor in the development and maintenance of addiction [2–4], across people with substance use disorders (SUD) [5] and laboratory animals exposed to drugs of abuse [6,7], but the specific contributing mechanisms remain unknown. Decision-making biases in addiction may be due to disruption in distinct components of learning, such as error encoding or value updating, or subjective preferences that are not readily observable in coarse behavioral performance measures. The nascent field of *computational psychiatry* applies formal models to understand the precise mechanisms (or "failure modes") that give rise to pathological

[*]**Correspondence**: Anna B. Konova PhD, Department of Psychiatry, University Behavioral Health Care, and the Brain Health Institute, Rutgers University—New Brunswick, 671 Hoes Lane West, Room D437, Piscataway, NJ 08855; tel.: 732-867-9720; anna.konova@rutgers.edu.

Conflict of interest statement
Nothing declared.

behavior in psychiatric conditions [8–10]. While there is no consensus on what qualifies as computational psychiatry, here we take this term to mean a mathematically rigorous understanding of the latent drivers of behavior. Findings from theory-driven computational psychiatry [11] suggest models that focus on algorithmic processes of value-based decision-making (Box 1) are well-suited to identify the specific components of reinforcement learning and decision-making that characterize SUD. This is exciting as such mechanistic research can bridge the behavioral manifestations of SUD with underlying neurobiology, providing fertile ground for cross-species translation [12–16]. Computational theoretical models thus hold promise as tools to provide additional mechanistic insight into SUD diagnosis and prognosis, and to help guide personalized treatments based on the latent variables governing individual behavior.

Here, we review recent theory-driven computational psychiatry studies of SUD primarily conducted with human subjects, highlighting the ways in which these studies have extended and refined our understanding of value-based decision-making processes in addiction. We focus on two key objectives of this work: to identify deviations from health (via case-control comparisons), and to map specific SUD symptoms and clinically-relevant states onto specific model variables—the latter aimed at moving closer to understanding the most defining yet most elusive aspect of the disorder: its dynamic, cyclical course. We conclude by outlining two directions for future research. We propose that a holistic approach that expands the typical parameter space examined within the same individual, and the duration of observation, may better serve these critical objectives and significantly enhance the clinical impact of computational psychiatry for addiction applications.

Deviation from health as indication of psychopathology: diagnostic differences between addicted and healthy individuals

SUD is a chronic, relapsing disorder characterized by repeated periods of drug craving, intoxication, bingeing, and withdrawal [17]. Drug use is maintained despite harmful consequences. The reinforcing and addictive effects of drugs center on the brain's reward (or "valuation" [18]) circuit. At the core of this circuit lie the dopaminergic pathways originating from the midbrain (ventral tegmental area and substantia nigra) and projecting onto the striatum and prefrontal cortex (orbitofrontal and ventromedial prefrontal cortex in particular). Dopaminergic circuits are intrinsic to reinforcement learning [19,20]. Decades of work in animal models suggests excessive stimulation of these circuits by drugs of abuse leads to an over-selection of drug-related actions at the expense of other adaptive behavior [2–4,6,7]. Early functional and molecular imaging work in humans also suggested abnormalities in dopaminergic function [21–23], potentially underlying abnormal value-based processes in SUD [5,24,25], but only recently have reinforcement learning (RL) mechanisms been dissected using formal modeling approaches.

Simple reinforcement learning

Combining functional brain imaging with computational modeling of choice behavior on simple ("model-free") RL paradigms, initial studies tested the theoretical assumption chronic drug users have deficits in value updating (Box 1), impeding learning from (non-

drug) reward and punishment outcomes. Contrary to theory, this research revealed minimal differences in fitted learning rates ([26–31], cf. [32]), and mixed evidence for reduced reward prediction error encoding in dopaminergic targets [28,33,34], with many finding no differences at all [26,27,29,31], in people with SUD across drug classes (nicotine, alcohol, stimulants, opioids) compared to healthy individuals. Further, under certain conditions, some users actually showed *increased* learning from punishment [32] and punishment prediction errors [34], while drugs with effects on dopamine administered acutely either normalized (rather than exacerbated) deviant learning phenotypes [32] or had no measurable impact on error encoding [33,34]. These data, together with subtle differences in quantitative measures of e.g., choice "stickiness" [30,32], strategic exploration [35], decision policy [36], and "transfer" of learning signals within frontostriatal circuits [27], hint that adaptations in other subprocesses of decision-making, or in the interplay between them, are involved in SUD.

Model-based/model-free reinforcement learning

To address these questions, recent work has leveraged additional tasks/models. Compulsive behavior in addiction is long thought to arise from a shift toward habitual and away from goal-directed behavioral control [37], a hypothesis that has found empirical support in some [38,39], though not all [40–42], studies in humans using outcome devaluation tests that can arbitrate between these controllers. Computationally, habitual and goal-directed control can be mapped onto distinct mechanisms: a "model-free" system that reflexively learns action-outcome contingencies and a "model-based" system requiring knowledge of the task structure, respectively (Box 1). Research examining these competing algorithms using sequential "two-stage" decision-making tasks has found evidence consistent with an imbalance in model-based vs. model-free learning in SUD across drug classes [43–46] that emerges only with chronic use [47]. Rather than overreliance on model-free RL, however, as might be expected from a habit account of addiction, this imbalance appears to stem from reduced model-based RL [43,45]. More directly, a computational re-analysis of the data in Ersche *et al.* [39] on a classic devaluation paradigm indicated that the tendency toward forming habits in stimulant users cannot be explained by model-free RL processes [48] (instead, increased 'reinforcement sensitivity'—a.k.a. inverse temperature—better accounted for users' behavior).

Taken together, computational approaches have permitted formal testing of theories of addiction. So far, the work on RL mechanisms reviewed here, and previously for alcohol [49], does not suggest the type of abnormality found in studies using coarser measures bears out in model-derived measures, as people with SUD do not appear to have reduced learning rates or reduced prediction error signaling relative to their healthy counterparts. This work is also beginning to shed light on related theories of habit learning, further showing that while SUD is associated with an imbalance in goal-directed vs. habitual control, this imbalance may not stem from differences in iterative learning from prediction errors as in model-free RL. This raises intriguing questions about what is at the core of observed behavioral biases in addiction. One possibility is that there is a complex interaction between internal drivers of this behavior, which may be missed by focusing on a single task/parameter and timepoint.

Economic choice and valuation

In support of this idea, parallel computational neuroeconomic studies find vast differences in people's "preferences", e.g., for delayed and risky—probabilistic—reward on paradigms that do not entail an explicit learning component (Box 1). Increased discount rates, or the rate at which the value of delayed reward diminishes with time to its delivery, have been reliably observed across SUD [50,51], and may stem from a similar latent decision process as model-based RL [52]. However, these variables are seldom measured together in the same individual. Similarly, risk preferences interact with RL processes [53], nonlinearly scaling prediction errors, and with discounting behavior [54]. Loss aversion, the idiosyncratic sensitivity to gains vs. losses, may further modulate learning, possibly accounting for some of the known asymmetry in positive and negative reward prediction error on choice. Importantly, studies applying formal economic models to quantify these preferences in addiction have found that, in aggregate, people with SUD have differential probability weighting [55], and are more risk tolerant [56] and less loss averse [57], than healthy individuals. To capture separable dimensions of value-based processes, and to more precisely map the resultant latent factors to SUD, we propose that a multidimensional examination of decision-making will be required, as illustrated in Figure 1.

Initial efforts to quantify multidimensional drivers of behavior in SUD took advantage of the Iowa Gambling Task. This complex decision-making task, widely used in the SUD literature, taps into both learning mechanisms and preferences (though these are not completely separately identifiable in the task). Computationally-informed analyses revealed poor learning on the IGT (captured by reduced average choice probabilities of higher reward-yielding options) was explained by reduced loss aversion in opioid users, increased risk tolerance in stimulant users, and by both alongside increased recency bias and reduced choice consistency in marijuana users [58,59]. In addition to providing initial support for interactions between value-based processes in SUD, this research also highlights previously unappreciated heterogeneity within SUD. Broadening the space of model parameters/tasks examined in the same individual, using the type of computational fingerprint approach we advocate, could provide a more detailed assessment of drug-specific effects [60] and a clearer mapping to clinical subtypes based on individual-level biological and clinical characteristics [14,61].

Such computational fingerprinting could take the form of a factor analysis to find lower-order dimensions, or principal components, in a space of model parameters, or via joint modeling of these parameters within an individual. This "fingerprint" could also be monitored across time to capture addiction-relevant transitions as discussed below and illustrated in Figure 1B and C, though we note this will require combining dimension reduction methods with complex trajectory analyses such as multidimensional scaling latent class/growth curve modeling.

Capturing addiction dynamics: using computational models to understand within-person variability, symptom expression, prognosis, and treatment

Addiction is not static, and indeed, it can be said that understanding addiction's longitudinal course is to understand addiction itself. The "addiction cycle" has been described as having three stages: preoccupation-anticipation, bingeing-intoxication, and withdrawal-negative affect [22,62–64]. These stages are likely associated with distinct value-based processes. Although no research to date has identified the algorithmic mechanisms that underlie the transition between each stage, initial work has advanced our understanding of the computational correlates of abstinence and withdrawal as well as features of preoccupation, namely craving. This more dynamic way of conceptualizing SUD holds great promise for realizing the clinical utility of computational psychiatry for addiction applications. In our proposed framework (Figure 1), the evolution of the computational fingerprint (multidimensional parameter space) can be used to identify critical periods when relapse vulnerability or treatment need is highest.

Abstinence and withdrawal

The most basic clinically-relevant transition is that between abstinence and use. Short-term abstinence, typically associated with aversive withdrawal states, has been associated with different RL mechanisms. Unsated vs. sated smokers were found to have reduced learning rates in the context of positive outcomes but enhanced learning from punishment [65], paralleling earlier observations of reduced prediction error encoding in striatum in this group [26]. However, others have observed more diffuse effects of nicotine abstinence [66]. Similarly, recently abstinent stimulant users, relative those with recent use, were found to have both selectively increased positive learning rates and heightened neural positive reward prediction errors [67], but reduced electrocortical signatures of both positive and negative reward prediction errors [68]. Finally, alcohol users with shorter abstinence durations exhibited more model-based/model-free imbalance [46]. Though mostly cross-sectional, these studies suggest computational measures may be used to dissect the specific mechanisms associated with abstinence/withdrawal states.

Craving

The preoccupation-anticipation stage of the addiction cycle is defined by intense subjective desire for the drug. Though it remains an open question what exactly craving "is", its importance in the maintenance of addiction cannot be overstated. In RL studies, craving has been associated with heightened frontostriatal encoding of prediction errors in drug deprived users [26,67]. Similarly, prediction error encoding in striatum was higher in smokers told there was nicotine in a smoked cigarette prior to an RL task (vs. when told there was no nicotine, and despite both cigarettes having nicotine) [69], an effect of drug expectation mediated by changes in insular activity and subjective craving [70]. No consistent relationship has been observed between craving and economic choice [57]. Indeed, craving appears to be an independent time-varying predictor of drug reuse when assessed alongside such measures [56]. More recently, computationally-informed conceptualizations of craving itself have been proposed, in which craving is defined as a

time- and attribute similarity-dependent multiplicative weight on value [71] or as a Bayesian process of hyper-precise prior estimates of interoceptive experience [72, 73]. However, key predictions of these models remain untested in SUD.

One important limitation is that almost all of the reviewed studies focus on non-drug reward. Particularly in the context of assessing craving (and arguably any theory of addiction), it is critical to test model predictions distinguishing between reward types (drug vs. non-drug) [42, 74]. Such studies may help answer questions about whether SUD is characterized by disrupted value-based processes broadly, or whether behavioral phenotypes are drug-stimulus specific and shifting across time as craving emerges.

Clinically-relevant transitions and treatment tailoring

Although addiction is defined by its longitudinal course, there has been a dearth of computationally-informed longitudinal research. At the chronic stages of SUD, the goal is to predict and hopefully prevent transitions within the addiction cycle (sustained abstinence or craving/withdrawal→drug use). People motivated to abstain, such as those initiating treatment, represent a clinically-important subgroup as well as one in which transitions are likely to occur on relatively short timescales (weeks to months). This provides an opportunity to address key questions about dynamic value-based processes. For example, such prior studies have found reduced model-based RL after detoxification predicted prospective 12-month alcohol relapse, though in combination with positive expectations about the reinforcing effects of alcohol [75].

Using a more temporally dense data collection protocol, we recently sought to identify proximal predictors of reuse events in treatment-engaged opioid users [56]. We measured two types of economic risk preferences (risk tolerance and ambiguity tolerance) repeatedly over 7 months and up to 15 times per person. We found that only ambiguity tolerance was associated with increased odds of prospective opioid use week-to-week. However, in aggregate, no significant differences in ambiguity tolerance were observed between opioid users and healthy controls, while opioid users were more risk tolerant regardless of reuse risk status. This suggests that even conceptually related value-based parameters may have distinct timecourses that convey distinct clinical information, further arguing for multidimensional assessment of this behavior.

There is strong theoretical impetus for treatments targeting value-based decision-making in addiction [76,77]. The identification and continuous monitoring of multidimensional computational fingerprints will be key for tailoring such interventions to the particular set of value-based processes at play for a given individual, at a given timepoint. Notably, elements of this proposed approach are already being tested in a new landmark study of computationally-informed behavior monitoring in SUD [12,78].

Of note, this initiative includes plans for cross-species work. This is critical as, in addition to permitting more precise investigation of neural circuits, animal models will prove particularly useful in addressing longer-timescale (developmental and lifespan) trajectories of value-based contributions to addiction that are impractical if not impossible to capture in humans. Emerging findings already support the utility of computational approaches for

interrogating RL mechanisms that are differentially altered preceding [79] vs. following [16,80,81] initiation of drug self-administration in rats, though as eluded to earlier these efforts may also be bolstered by re-analysis of existing data.

Conclusion and future directions

Computational psychiatry has garnered considerable attention in recent years but enthusiasm for its presumed clinical utility is rightly tempered [82]. Here, we review the promise of this approach for addiction applications. While computationally-informed studies have produced novel explanatory insights about value-based processes in addiction that help to refine long-held theoretical accounts, we also identified two directions for future research that could significantly enhance the clinical translational potential of this approach.

First, we emphasize the importance of multidimensional assessment. Currently, multiple computational mechanisms are rarely assessed within the same individual, precluding identification of shared and distinct latent constructs underpinned by distinct neural substrates. SUD is extremely heterogeneous, including differences in the pharmacological actions of different drugs, patterns of use, symptom phenomenology, and availability of adequate treatments. A multidimensional assessment of value-based decision-making could provide the needed precision for mapping computational mechanisms to heterogeneous clinical profiles. Ultimately, this might allow for identification of person-specific combinations of model parameters underlying different disease mechanisms.

Critically, such quantifiable computational fingerprints should be examined longitudinally. The most defining feature of addiction is its cyclic course. This has not been adequately captured in prior work. Most research remains cross-sectional, emphasizing between-person differences. We propose that a “holistic”, longitudinal and multidimensional examination of value-based processes within-person, a type of *dynamic* computational fingerprint, will provide a more complete understanding of addiction as well as aid in the development of better tailored and timed interventions.

Acknowledgments

The authors acknowledge funding from the Brain & Behavior Research Foundation (BBRF NARSAD Grant #25387), Busch Biomedical Research Program, and NIH/NIDA (DA043676). Special thanks to the Addiction and Decision Neuroscience Laboratory members, Silvia Lopez-Guzman, and Paul W. Glimcher for helpful discussions.

References

Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest

•• of outstanding interest

1. Rangel A, Camerer C, Montague PR: A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci2008, 9:545–556. [PubMed: 18545266]
2. Redish AD, Jensen S, Johnson A: A unified framework for addiction: vulnerabilities in the decision process. Behav Brain Sci2008, 31:415–437; discussion 437–487. [PubMed: 18662461]

3. Redish AD: Addiction as a computational process gone awry. Science2004, 306:1944–1947. [PubMed: 15591205]
4. Huys QJ, Tobler PN, Hasler G, Flagel SB: The role of learning-related dopamine signals in addiction vulnerability. Prog Brain Res2014, 211:31–77. [PubMed: 24968776]
5. Bickel WK, Mellis AM, Snider SE, Athamneh LN, Stein JS, Pope DA: 21st century neurobehavioral theories of decision making in addiction: Review and evaluation. Pharmacol Biochem Behav2018, 164:4–21. [PubMed: 28942119]
6. Schoenbaum G, Roesch MR, Stalnaker TA: Orbitofrontal cortex, decision-making and drug addiction. Trends Neurosci2006, 29:116–124. [PubMed: 16406092]
7. Ahmed SH: Individual decision-making in the causal pathway to addiction: contributions and limitations of rodent models. Pharmacol Biochem Behav2018, 164:22–31. [PubMed: 28709784]
8. Montague PR, Dolan RJ, Friston KJ, Dayan P: Computational psychiatry. Trends Cogn Sci2012, 16:72–80. [PubMed: 22177032]
9. Wang XJ, Krystal JH: Computational psychiatry. Neuron2014, 84:638–654. [PubMed: 25442941]
10••. Huys QJ, Maia TV, Frank MJ: Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci2016, 19:404–413. [PubMed: 26906507] Seminal review defining the two branches of computational psychiatry, data-driven and theory-driven, with an overview of their respective contributions for precision psychiatry and mechanistic understanding of disease states.
11. Maia TV, Huys QJM, Frank MJ: Theory-Based Computational Psychiatry. Biol Psychiatry2017, 82:382–384. [PubMed: 28838466]
12. Liu S, Dolan RJ, Heinz A: Translation of Computational Psychiatry in the Context of Addiction. JAMA Psychiatry2020.
13. Belin-Rauscent A, Fouyssac M, Bonci A, Belin D: How Preclinical Models Evolved to Resemble the Diagnostic Criteria of Drug Addiction. Biol Psychiatry2016, 79:39–46. [PubMed: 25747744]
14. Sweis BM, Thomas MJ, Redish AD: Beyond simple tests of value: measuring addiction as a heterogeneous disease of computation-specific valuation processes. Learn Mem2018, 25:501–512. [PubMed: 30115772]
15. Groman SM: The Neurobiology of Impulsive Decision-Making and Reinforcement Learning in Nonhuman Animals. Curr Top Behav Neurosci2020.
16. Groman SM: Investigating the computational underpinnings of addiction. Neuropsychopharmacology2019, 44:2149–2150. [PubMed: 31076651]
17. DSM5: Diagnostic and statistical manual of mental disordersedn 5th. Arlington, VA: American Psychiatric Association; 2013.
18. Bartra O, McGuire JT, Kable JW: The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage2013, 76:412–427. [PubMed: 23507394]
19. Bayer HM, Glimcher PW: Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron2005, 47:129–141. [PubMed: 15996553]
20. Schultz W, Dayan P, Montague PR: A neural substrate of prediction and reward. Science1997, 275:1593–1599. [PubMed: 9054347]
21. Trifilieff P, Martinez D: Blunted dopamine release as a biomarker for vulnerability for substance use disorders. Biol Psychiatry2014, 76:4–5. [PubMed: 24925889]
22. Goldstein RZ, Volkow ND: Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. Nat Rev Neurosci2011, 12:652–669. [PubMed: 22011681]
23. Volkow ND, Koob GF, McLellan AT: Neurobiologic Advances from the Brain Disease Model of Addiction. N Engl J Med2016, 374:363–371. [PubMed: 26816013]
24. Luijten M, Schellekens AF, Kuhn S, Machielse MW, Sescousse G: Disruption of Reward Processing in Addiction : An Image-Based Meta-analysis of Functional Magnetic Resonance Imaging Studies. JAMA Psychiatry2017, 74:387–398. [PubMed: 28146248]
25. Konova AB, Goldstein RZ: Role of the value circuit in addiction and addiction treatment. In The Wiley-Blackwell handbook on the neuroscience of addiction. Edited by Wiley-Blackwell: Wiley-Blackwell; 2015.

26. Chiu PH, Lohrenz TM, Montague PR: Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. Nat Neurosci2008, 11:514–520. [PubMed: 18311134]
27. Park SQ, Kahnt T, Beck A, Cohen MX, Dolan RJ, Wrase J, Heinz A: Prefrontal cortex fails to learn from reward prediction errors in alcohol dependence. J Neurosci2010, 30:7749–7753. [PubMed: 20519550]
28. Tanabe J, Reynolds J, Krmpotich T, Claus E, Thompson LL, Du YP, Banich MT: Reduced neural tracking of prediction error in substance-dependent individuals. Am J Psychiatry2013, 170:1356–1363. [PubMed: 23897123]
29. Gradin VB, Baldacchino A, Balfour D, Matthews K, Steele JD: Abnormal brain activity during a reward and loss task in opiate-dependent patients receiving methadone maintenance therapy. Neuropsychopharmacology2014, 39:885–894. [PubMed: 24132052]
30. Myers CE, Sheynin J, Balsdon T, Luzardo A, Beck KD, Hogarth L, Haber P, Moustafa AA: Probabilistic reward- and punishment-based learning in opioid addiction: Experimental and computational data. Behav Brain Res2016, 296:240–248. [PubMed: 26381438]
31. Deserno L, Beck A, Huys QJ, Lorenz RC, Buchert R, Buchholz HG, Plotkin M, Kumakara Y, Cumming P, Heinze HJ, et al.: Chronic alcohol intake abolishes the relationship between dopamine synthesis capacity and learning signals in the ventral striatum. Eur J Neurosci2015, 41:477–486. [PubMed: 25546072]
32. Kanen JW, Ersche KD, Fineberg NA, Robbins TW, Cardinal RN: Computational modelling reveals contrasting effects on reinforcement learning and cognitive flexibility in stimulant use disorder and obsessive-compulsive disorder: remediating effects of dopaminergic D2/3 receptor agents. Psychopharmacology (Berl)2019, 236:2337–2358. [PubMed: 31324936]
33. Rose EJ, Ross TJ, Salmeron BJ, Lee M, Shakleya DM, Huestis M, Stein EA: Chronic exposure to nicotine is associated with reduced reward-related activity in the striatum but not the midbrain. Biol Psychiatry2012, 71:206–213. [PubMed: 22032832]
34. Rose EJ, Salmeron BJ, Ross TJ, Waltz J, Schweitzer JB, McClure SM, Stein EA: Temporal difference error prediction signal dysregulation in cocaine dependence. Neuropsychopharmacology2014, 39:1732–1742. [PubMed: 24569319]
35. Morris LS, Baek K, Kundu P, Harrison NA, Frank MJ, Voon V: Biases in the Explore-Exploit Tradeoff in Addictions: The Role of Avoidance of Uncertainty. Neuropsychopharmacology2016, 41:940–948. [PubMed: 26174598]
36. Harle KM, Zhang S, Schiff M, Mackey S, Paulus MP, Yu AJ: Altered Statistical Learning and Decision-Making in Methamphetamine Dependence: Evidence from a Two-Armed Bandit Task. Front Psychol2015, 6:1910. [PubMed: 26733906]
37. Luscher C, Robbins TW, Everitt BJ: The transition to compulsion in addiction. Nat Rev Neurosci2020, 21:247–263. [PubMed: 32231315]
38. Sjoerds Z, de Wit S, van den Brink W, Robbins TW, Beekman AT, Penninx BW, Veltman DJ: Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. Transl Psychiatry2013, 3:e337. [PubMed: 24346135]
39••. Ersche KD, Gillan CM, Jones PS, Williams GB, Ward LH, Luijten M, de Wit S, Sahakian BJ, Bullmore ET, Robbins TW: Carrots and sticks fail to change behavior in cocaine addiction. Science2016, 352:1468–1471. [PubMed: 27313048] This study found that overtraining with positive reinforcement led to more habitual behavior in cocaine users, as confirmed by a devaluation test. By contrast, overtraining with punishment had no such effect. A follow-up computational dissection of these data (Lim et al. 2019) however revealed model-free RL did not fully explain this tendency in cocaine users, highlighting the need for alternative models.
40. Luijten M, Gillan CM, de Wit S, Franken IHA, Robbins TW, Ersche KD: Goal-Directed and Habitual Control in Smokers. Nicotine Tob Res2020, 22:188–195. [PubMed: 30768206]
41. van Timmeren T, Quail SL, Balleine BW, Geurts DEM, Goudriaan AE, van Holst RJ: Intact corticostriatal control of goal-directed action in Alcohol Use Disorder: a Pavlovian-to-instrumental transfer and outcome-devaluation study. Sci Rep2020, 10:4949. [PubMed: 32188908]
42. Hogarth L: Addiction is driven by excessive goal-directed drug choice under negative affect: translational critique of habit and compulsion theory. Neuropsychopharmacology2020, 45:720–735. [PubMed: 31905368]

43. Sebold M, Deserno L, Nebe S, Schad DJ, Garbusow M, Hagele C, Keller J, Junger E, Kathmann N, Smolka MN, et al.: Model-based and model-free decisions in alcohol dependence. Neuropsychobiology2014, 70:122–131. [PubMed: 25359492]
44. Voon V, Derbyshire K, Ruck C, Irvine MA, Worbe Y, Enander J, Schreiber LR, Gillan C, Fineberg NA, Sahakian BJ, et al.: Disorders of compulsivity: a common bias towards learning habits. Mol Psychiatry2015, 20:345–352. [PubMed: 24840709]
45. Reiter AM, Deserno L, Kallert T, Heinze HJ, Heinz A, Schlagenhauf F: Behavioral and Neural Signatures of Reduced Updating of Alternative Options in Alcohol-Dependent Patients during Flexible Decision-Making. J Neurosci2016, 36:10935–10948. [PubMed: 27798176]
46. Donamayor N, Strelchuk D, Baek K, Banca P, Voon V: The involuntary nature of binge drinking: goal directedness and awareness of intention. Addict Biol2018, 23:515–526. [PubMed: 28419776]
47. Nebe S, Kroemer NB, Schad DJ, Bernhardt N, Sebold M, Muller DK, Scholl L, Kuitunen-Paul S, Heinz A, Rapp MA, et al.: No association of goal-directed and habitual control with alcohol consumption in young adults. Addict Biol2018, 23:379–393. [PubMed: 28111829]
48. Lim TV, Cardinal RN, Savulich G, Jones PS, Moustafa AA, Robbins TW, Ersche KD: Impairments in reinforcement learning do not explain enhanced habit formation in cocaine use disorder. Psychopharmacology (Berl)2019, 236:2359–2371. [PubMed: 31372665]
49. Huys QJM, Deserno L, Obermayer K, Schlagenhauf F, Heinz A: Model-Free Temporal-Difference Learning and Dopamine in Alcohol Dependence: Examining Concepts From Theory and Animals in Human Imaging. Biol Psychiatry Cogn Neurosci Neuroimaging2016, 1:401–410. [PubMed: 29560869]
50. MacKillop J, Amlung MT, Few LR, Ray LA, Sweet LH, Munafo MR: Delayed reward discounting and addictive behavior: a meta-analysis. Psychopharmacology (Berl)2011, 216:305–321. [PubMed: 21373791]
51. Amlung M, Vedelago L, Acker J, Balodis I, MacKillop J: Steep delay discounting and addictive behavior: a meta-analysis of continuous associations. Addiction2017, 112:51–62.
52. Hunter LE, Bornstein AM, Hartley CA: A common deliberative process underlies model-based planning and patient intertemporal choice. bioRxiv2018:499707.
53. Niv Y, Edlund JA, Dayan P, O'Doherty JP: Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. J Neurosci2012, 32:551–562. [PubMed: 22238090]
54. Lopez-Guzman S, Konova AB, Glimcher PW: Computational psychiatry of impulsivity and risk: how risk and time preferences interact in health and disease. Philos Trans R Soc Lond B Biol Sci2019, 374:20180135. [PubMed: 30966919]
55. Voon V, Morris LS, Irvine MA, Ruck C, Worbe Y, Derbyshire K, Rankov V, Schreiber LR, Odlaug BL, Harrison NA, et al.: Risk-taking in disorders of natural and drug rewards: neural correlates and effects of probability, valence, and magnitude. Neuropsychopharmacology2015, 40:804–812. [PubMed: 25270821]
56••. Konova AB, Lopez-Guzman S, Urmanche A, Ross S, Louie K, Rotrosen J, Glimcher PW: Computational Markers of Risky Decision-making for Identification of Temporal Windows of Vulnerability to Opioid Use in a Real-world Clinical Setting. JAMA Psychiatry2019.This study employed a serial longitudinal design to capture fluctuations in value-based decision-making that mapped onto dynamic clinical phenomena in opioid addiction, laying the groundwork for a new framework for the computational psychiatry study of the cycle of addiction within-person.
57. Genauck A, Quester S, Wustenberg T, Morsen C, Heinz A, Romanczuk-Seiferth N: Reduced loss aversion in pathological gambling and alcohol dependence is associated with differential alterations in amygdala and prefrontal functioning. Sci Rep2017, 7:16306. [PubMed: 29176580]
58•. Ahn WY, Vasilev G, Lee SH, Busemeyer JR, Kruschke JK, Bechara A, Vassileva J: Decision-making in stimulant and opiate addicts in protracted abstinence: evidence from computational modeling with pure users. Front Psychol2014, 5:849. [PubMed: 25161631] Study that employs a computational dissection of behavior on the classic IGT task, revealing complex interactions between stable preferences and RL, and distinct subprocesses associated with learning deficits in different types of substance users (opioid vs. stimulant).

59. Fridberg DJ, Queller S, Ahn WY, Kim W, Bishara AJ, Busemeyer JR, Porrino L, Stout JC: Cognitive Mechanisms Underlying Risky Decision-Making in Chronic Cannabis Users. J Math Psychol2010, 54:28–38. [PubMed: 20419064]

60. Badiani A, Caprioli D, De Pirro S: Opposite environmental gating of the experienced utility ('liking') and decision utility ('wanting') of heroin versus cocaine in animals and humans: implications for computational neuroscience. Psychopharmacology (Berl)2019, 236:2451–2471. [PubMed: 31289884]

61. Feczko E, Miranda-Dominguez O, Marr M, Graham AM, Nigg JT, Fair DA: The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes. Trends Cogn Sci2019, 23:584–601. [PubMed: 31153774]

62. Koob GF, Le Moal M: Drug abuse: hedonic homeostatic dysregulation. Science1997, 278:52–58. [PubMed: 9311926]

63. Goldstein RZ, Volkow ND: Drug addiction and its underlying neurobiological basis: neuroimaging evidence for the involvement of the frontal cortex. Am J Psychiatry2002, 159:1642–1652. [PubMed: 12359667]

64. Koob GF, Volkow ND: Neurocircuitry of addiction. Neuropsychopharmacology2010, 35:217–238. [PubMed: 19710631]

65•. Baker TE, Zeighami Y, Dagher A, Holroyd CB: Smoking Decisions: Altered Reinforcement Learning Signals Induced by Nicotine State. Nicotine Tob Res2020, 22:164–171. [PubMed: 29982681] This study examined how different smoking states (abstinence, ad lib consumption) impacted model-free RL within-person, finding that short-term abstinence reduced learning rates in the context of positive outcomes but enhanced learning from punishment.

66. Lesage E, Aronson SE, Sutherland MT, Ross TJ, Salmeron BJ, Stein EA: Neural Signatures of Cognitive Flexibility and Reward Sensitivity Following Nicotinic Receptor Stimulation in Dependent Smokers: A Randomized Trial. JAMA Psychiatry2017, 74:632–640. [PubMed: 28403383]

67•. Wang JM, Zhu L, Brown VM, De La Garza R 2nd, Newton T, King-Casas B, Chiu PH: In Cocaine Dependence, Neural Prediction Errors During Loss Avoidance Are Increased With Cocaine Deprivation and Predict Drug Use. Biol Psychiatry Cogn Neurosci Neuroimaging2019, 4:291–299. [PubMed: 30297162] Detailed study highlighting the influence of state within a person (drug use or deprivation) on neural prediction error signals and model-derived learning rates, providing evidence for the biological underpinnings of computational processes that change across addiction states.

68. Parvaz MA, Konova AB, Proudfit GH, Dunning JP, Malaker P, Moeller SJ, Maloney T, Alia-Klein N, Goldstein RZ: Impaired neural response to negative prediction errors in cocaine addiction. JNeurosci2015, 35:1872–1879. [PubMed: 25653348]

69. Gu X, Lohrenz T, Salas R, Baldwin PR, Soltani A, Kirk U, Cinciripini PM, Montague PR: Belief about nicotine selectively modulates value and reward prediction error signals in smokers. Proc Natl Acad Sci U S A2015, 112:2539–2544. [PubMed: 25605923]

70. Gu X, Lohrenz T, Salas R, Baldwin PR, Soltani A, Kirk U, Cinciripini PM, Montague PR: Belief about Nicotine Modulates Subjective Craving and Insula Activity in Deprived Smokers. Front Psychiatry2016, 7:126. [PubMed: 27468271]

71. Konova AB, Louie K, Glimcher PW: The computational form of craving is a selective multiplication of economic value. Proc Natl Acad Sci U S A2018, 115:4122–4127. [PubMed: 29610355]

72•. Gu X, Filbey F: A Bayesian Observer Model of Drug Craving. JAMA Psychiatry2017, 74:419–420. [PubMed: 28199454] This paper presents a Bayesian model of craving as a fluid symptom that stems from hyper-precise internal priors about interoceptive experience that can evolve over time, in the approach to and recovery from drug use.

73. Gu X: Incubation of craving: a Bayesian account. Neuropsychopharmacology2018, 43:2337–2339. [PubMed: 29930384]

74. Hogarth L, Field M: Relative expected value of drugs versus competing rewards underpins vulnerability to and recovery from addiction. Behav Brain Res2020:112815. [PubMed: 32707138]

75•. Sebold M, Nebe S, Garbusow M, Guggenmos M, Schad DJ, Beck A, Kuitunen-Paul S, Sommer C, Frank R, Neu P, et al.: When Habits Are Dangerous: Alcohol Expectancies and Habitual Decision Making Predict Relapse in Alcohol Dependence. Biol Psychiatry2017, 82:847–856. [PubMed: 28673442] This study examined whether reduced use of model-based RL strategies, previously found to distinguish patients and controls, predicted the transition to alcohol relapse. This was only the case when considering individual alcohol expectancies, suggesting additional and interacting processes are at play in relapse vulnerability.

76. Heinz A, Deserno L, Zimmermann US, Smolka MN, Beck A, Schlagenhauf F: Targeted intervention: Computational approaches to elucidate and predict relapse in alcoholism. Neuroimage2017, 151:33–44. [PubMed: 27480622]

77. Verdejo-Garcia A, Alcazar-Corcoles MA, Albein-Urios N: Neuropsychological Interventions for Decision-Making in Addiction: a Systematic Review. Neuropsychol Rev2019, 29:79–92. [PubMed: 30259388]

78•. Heinz A, Kiefer F, Smolka MN, Endrass T, Beste C, Beck A, Liu S, Genauck A, Romund L, Banaschewski T, et al.: Addiction Research Consortium: Losing and regaining control over drug intake (ReCoDe)-From trajectories to mechanisms and interventions. Addict Biol2020, 25:e12866. [PubMed: 31859437] Planned landmark study spanning 12 years and aiming to dynamically characterize alcohol addiction computational phenotypes in humans and non-human animals.

79••. Groman SM, Massi B, Mathias SR, Lee D, Taylor JR: Model-Free and Model-Based Influences in Addiction-Related Behaviors. Biol Psychiatry2019, 85:936–945. [PubMed: 30737015] Preclinical study showing the promise of computational approaches for translational addiction neuroscience, finding that as in humans, chronic drug exposure is associated with reduced model-based RL while increased model-free RL precedes escalating addiction phenotypes.

80. Zhukovsky P, Puaud M, Jupp B, Sala-Bayo J, Alsio J, Xia J, Searle L, Morris Z, Sabir A, Giuliano C, et al.: Withdrawal from escalated cocaine self-administration impairs reversal learning by disrupting the effects of negative feedback on reward exploitation: a behavioral and computational analysis. Neuropsychopharmacology2019, 44:2163–2173. [PubMed: 30952156]

81. Groman SM, Hillmer AT, Liu H, Fowles K, Holden D, Morris ED, Lee D, Taylor JR: Dysregulation of Decision Making Related to Metabotropic Glutamate 5, but Not Midbrain D3, Receptor Availability Following Cocaine Self-administration in Rats. Biol Psychiatry2020.

82. Browning M, Carter CS, Chatham C, Den Ouden H, Gillan CM, Baker JT, Chekroud AM, Cools R, Dayan P, Gold J, et al.: Realizing the Clinical Potential of Computational Psychiatry: Report From the Banbury Center Meeting, February 2019. Biol Psychiatry2020, 88:e5–e10. [PubMed: 32113656]

Box 1.

Common models of reinforcement learning and decision-making in research on addiction and key model parameter definitions.

Simple Reinforcement Learning	**Key estimated parameters** α- Learning rate, rate at which past outcomes influence current choices
Standard learning model based on learning rate and prediction errors that are used to update action-outcome (or stimulus-outcome) associations	Q learning $\delta_t = R_t - Q_t$ $Q_{t+1} = Q_t + \alpha \cdot \delta_t$
Model-Based/Model-Free Reinforcement Learning	**Key estimated parameters** α- Learning rate, rate at which past outcomes influence current choices ω- Weight parameter to determine relative influence of MB vs. MF
Based on learning that is updated using a balance of previous prediction error from past choices and knowledge of the task structure with the available actions (a) at each state (s), and typically tested with "2-stage" tasks.	Model-Free (MF) $Q_{MF}(s_{i,t+1}, a_{i,t+1}) = Q_{MF}(s_{i,t}, a_{i,t}) + \alpha_i \cdot \delta_{i,t}$ $Q_{MF}(s_{1,t}, a_{1,t}) = Q_{MF}(s_{1,t}, a_{1,t}) + \alpha_1 \cdot \lambda\delta_{2,t}$ (where λ is an eligibility trace allowing outcome at 2nd stage to influence 1st stage choice) Model-Based (MB) $Q_{MB}(s_A, a_j) = P(s_B \mid s_A, a_j) \cdot max\, Q_{MF}(s_B, a) + P(s_C \mid s_A, a_j) \cdot max\, Q_{MF}(s_C, a)$ MB-MF balance $Q_{net}(s_A, a_j) = \omega \cdot Q_{MB}(s_A, a_j) + (1 - \omega) \cdot Q_{MF}(s_A, a_j)$
Economic Choice and Valuation *Discounting*. how temporal factors depreciate value when reward/gratification is delayed	**Key estimated parameters** κ- Discount rate, measure of attitude towards delayed rewards α- Risk tolerance, measure of attitude towards risky rewards β- Ambiguity tolerance, measure of attitude towards ambiguous rewards λ- Loss aversion, measure of avoidance of potential loss B- Sensitivity to losses and gains
Risk preference. how individual attitudes about known risk and ambiguity influence the value of choice options *Loss aversion*. the balance between individual gain and loss sensitivities	Hyperbolic discounting $U_{option} = \frac{v}{1 + \kappa D}$ Expected utility theory with only risk $U_{option} = p \cdot v^{\alpha}$ Expected utility theory with risk and ambiguity $U_{option} = \left(p - \beta\frac{A}{2}\right) \cdot v^{\alpha}$ Prospect theory $U_{option} = \pi(p_i) \cdot v(x_i)$ Loss aversion $\lambda = \frac{\lvert B_{loss} \rvert}{B_{gain}}$

Highlights

- Computational psychiatry holds promise for mechanistic discovery in addiction
- This approach captures latent factors driving behavioral differences from health
- Emerging support also for capturing variation defining addiction cycles and states
- Research needs to better account for the heterogeneous, dynamic nature of addiction
- Expanding the parameter space examined and duration of observation will be key

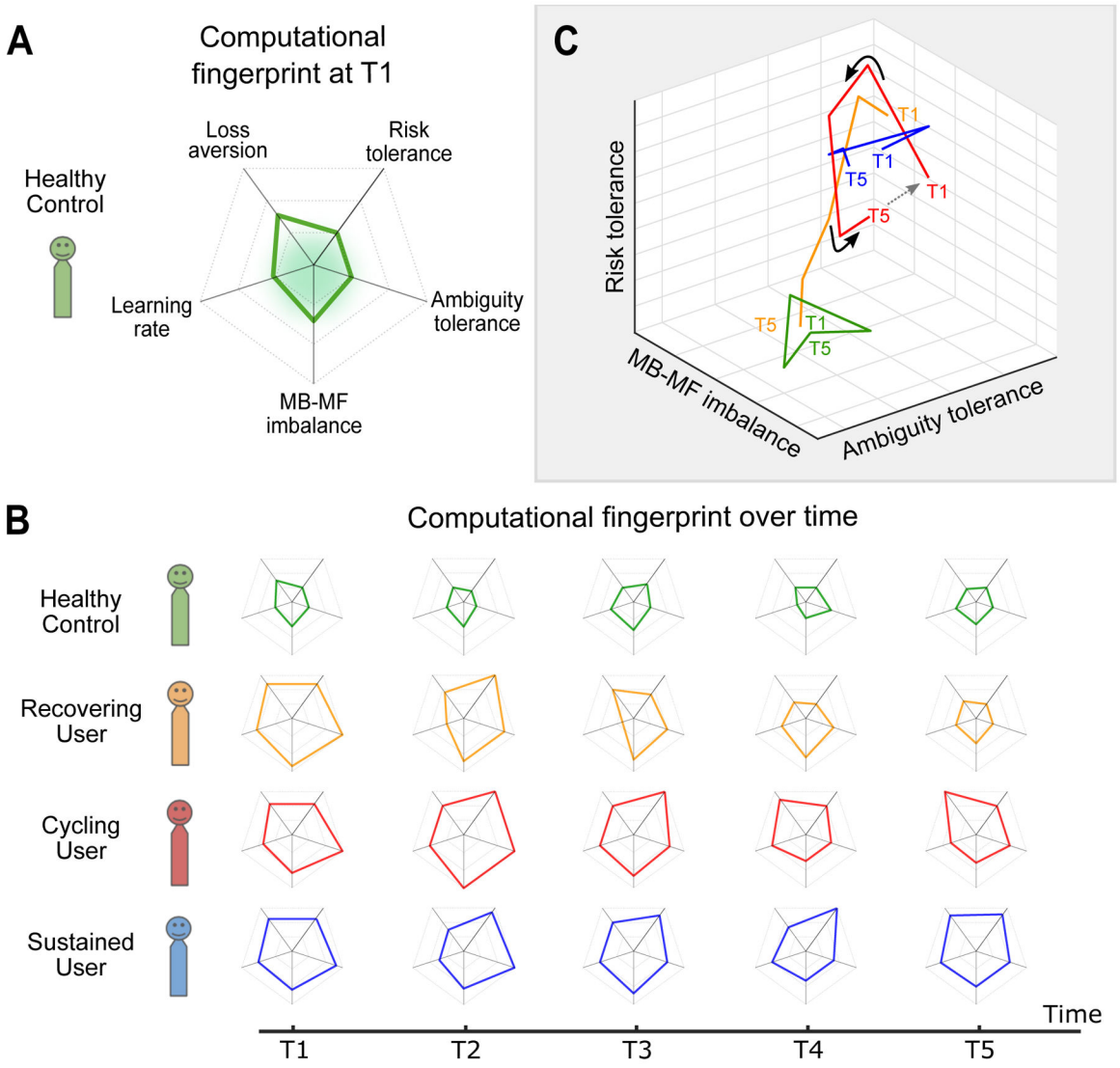

Figure 1.

Computational "fingerprinting" and dynamic characterization of addiction trajectories and transitions.

(A) The computational parameter space (a type of "computational fingerprint") of a healthy individual showing select value-based decision-making parameters reported in the reviewed studies as being altered in addiction. The green shaded area represents a "healthy norm".

(B) Fluctuations over time of the computational fingerprints for prognosis-based addiction classification, i.e. recovering, cycling (abstinence and relapse stages), and sustained use cases as compared to health (green).

(C) Evolution of the parameter space over time shown here for three components of the fingerprint for illustrative purposes (the full space may contain additional components). The example cases represent realistic trajectories/states: 1) healthy, shown to stay at the same multidimensional space over time; 2) sustained use, also shown to stay in the same space over time but to occupy a different one from health; 3) cycling use, shown to move away from an initial starting point and then start to return back to it. Here we also highlight at

what time points tailored treatment might be most efficacious (i.e. when individuals might be most susceptible to intervention strategies) designated by the solid arrows (→); and 4) recovering, also shown to move but in a single direction approaching health. Note that here "component" could be a single estimated parameter (as shown in the 3D plot), a single estimated parameter accounting for the influence of another parameter (e.g., risk-preference adjusted learning rate), or a principal component (dimension comprised of a combination of parameters).