



REVIEW ARTICLE

Habit, choice, and addiction

Y. Vandaele¹ and S. H. Ahmed^{2,3}

Addiction was suggested to emerge from the progressive dominance of habits over goal-directed behaviors. However, it is generally assumed that habits do not persist in choice settings. Therefore, it is unclear how drug habits may persist in real-world scenarios where this factor predominates. Here, we discuss the poor translational validity of the habit construct, which impedes our ability to determine its role in addiction. New evidence of habitual behavior in a drug choice setting are then described and discussed. Interestingly, habitual preference did not promote drug choice but instead favored abstinence. Here, we propose several clues to reconcile these unexpected results with the habit theory of addiction, and we highlight the need in experimental research to face the complexity of drug addicts' decision-making environments by investigating drug habits in the context of choice and in the presence of cues. On a theoretical level, we need to consider more complex frameworks, taking into account continuous interactions between goal-directed and habitual systems, and alternative decision-making models more representative of real-world conditions.

Neuropsychopharmacology (2021) 46:689–698; <https://doi.org/10.1038/s41386-020-00899-y>

INTRODUCTION

Tobacco, alcohol, and substance use disorders, which will be referred to as addiction in the present review, are all driven by a transition toward compulsive drug use characterized by a loss of control over drug intake, persistent drug use despite dreadful consequences, and frequent episodes of relapse. Among recreational users, only a subset ultimately lose control over drug use and develop an addiction. To explain this transition, several, often overlapping, theories have been proposed [1]. Among them, the influential but controversial habit theory of addiction posits that the transition to addiction emerges from the progressive development and dominance of drug habits over goal-directed control [2, 3]. Although drug habits appear omnipresent in any form of addiction, whether formation or expression of drug habits contribute to the transition to addiction remains a matter of debate.

The involvement of automatic processes in addiction was suggested 30 years ago in the seminal work of Tiffany [4]. Several diagnostic criteria for SUD are consistent with the concept of drug habit; notably, the persistence of drug use when it is no longer pleasurable and despite negative consequences, the high reactivity to drug-associated cues and context, and the fact that addictive behaviors appear out of voluntary control [1, 5, 6]. Habits are defined as automatic responses elicited by antecedent stimuli without deliberation or representation of the consequences of one's action. Because habits do not depend on the response–outcome association underlying goal-directed behavior, they are generally operationalized as an absence of goal-directed behavior; that is, actions not affected by a reduction of the outcome value and/or by a degradation of the response–outcome contingency are under habitual control (Box 1) [7, 8]. Although these tests typically answer a yes-or-no question, habit and goal-directed systems likely control behavior along a continuum, and

the balance between these two systems would be shifted toward habit in SUD.

However, the relation between drug use and habit remains controversial in humans, with mixed results and significant discrepancies [9, 10]. Furthermore, although the literature in rodents converges to show that drug exposure promotes habit, how drug habits favor further drug use and, ultimately, the transition to addiction remains unclear. In this review, we try to address this question by reviewing behavioral evidence supporting the habit theory of addiction in rodents and discussing important limitations, notably the absence of habit in choice settings. We then present new evidence of habitual behavior in a drug choice setting and propose several clues to explain our unexpected results in the light of the habit theory of addiction. We propose new perspectives on this theory that embrace the complexity of the decision-making environment of drug addicts and of interactions between decision-making processes.

DRUGS PROMOTE HABIT

A large number of studies in rodents show that drugs of abuse promote habit. Following drug self-administration training, drugs can be devalued using either sensory-specific satiety or CTA before responding for the drug is tested under extinction (Box 1). Using this procedure, it was shown that responding for ethanol [11–17], cocaine [18, 19], and nicotine [20, 21] becomes habitual after various length of training. In some studies, the transition to habit was faster for the drug compared to a nondrug reward suggesting stronger facilitation of habit formation for drug seeking [11, 13, 15, 18, 21]. Interestingly, studies in which rats are trained to self-administer cocaine or heroin in a seeking-taking schedule (e.g., heterogeneous chains; seeking RI30—taking FR1 on

¹Department of Psychiatry, Lausanne University Hospital, Lausanne, Switzerland; ²Institut des Maladies Neurodégénératives, Université de Bordeaux, Bordeaux, France and

³Institut des Maladies Neurodégénératives, CNRS, Bordeaux, France

Correspondence: Y. Vandaele (youna.vandaele@chuv.ch)

Received: 24 August 2020 Revised: 7 October 2020 Accepted: 19 October 2020

Published online: 9 November 2020

Box 1 Experimental tests of habitual control

In contrast to goal-directed behavior, habit does not depend on the current motivational value of the outcome and on the knowledge of a causal relationship between the response and the outcome. Thus, reducing the value of the outcome and/or the contingency between the response and the outcome does not affect habitual behavior but reduces responding under goal-directed control (Balleine and Dickinson [8]; Dickinson [41]; Dickinson and Balleine [7]).

Outcome devaluation: the value of a reward is typically reduced by sensory-specific satiety or by pairing the consumption of the reward with an injection of lithium chloride to induce conditioned taste aversion (CTA) (Adams and Dickinson 1981 [129]; Balleine and Dickinson [8]; Colwill 1993 [126]; Dickinson and Balleine [7]; Rescorla 1987 [128]). Responding for the devalued outcome is then tested under extinction and compared to a control condition in which the outcome is not devalued.

Contingency degradation: the contingency between the response and the outcome can be degraded by providing noncontingent delivery of one outcome, while maintaining another response–outcome association intact. For instance, one action (R1) is performed to obtain a reward (O1), while another action (R2) gives access to another reward (O2). During the test, one of the outcomes (i.e., O1) is delivered non contingently such that its delivery is equally probable following a response or not (that is, $p(O1/R1) = p(O1/\sim R1) = 0.5$). The contingency of this R1–O1 association is thus degraded. The alternative R2–O2 contingency remains intact. Goal-directed performance of the degraded reward should be reduced compared to the non-degraded alternative (Colwill 1993 [126]; Dickinson and Mulatero 1989 [127]). Conversely, insensitivity to this procedure indicates that performance is under habitual control.

separate levers) reveal that rats correctly encode the contingency between the seeking response, the taking response and the outcome, indicating that their behavior is under goal-directed control [22, 23]. However, it was also shown that the cocaine-seeking response becomes insensitive to extinction of the cocaine taking response following extended self-administration training, suggesting a shift to habitual control [24].

Numerous studies show that passive drug exposure is sufficient to promote habitual responding for nondrug rewards. For instance, while lever pressing for a solution of 20% sucrose remains under goal-directed control after 8 weeks of training, home-cage access to ethanol during instrumental training renders the behavior habitual [11]. Ethanol-induced facilitation of habitual responding for food was also found following chronic intermittent exposure to ethanol vapor [25]. Passive cocaine [26, 27] or amphetamine [28–30] exposure also rendered responding for a nondrug reward insensitive to devaluation by specific satiety or CTA. Interestingly, even limited post-training exposure to cocaine was sufficient to observe habitual responding for food rewards [31], a results not replicated with amphetamine [32]. Drug-induced facilitation of habit was also demonstrated in studies showing insensitivity to degradation of instrumental contingency (Box 1) following ethanol exposure [16] or repeated injections of cocaine [33]. However, two studies have found that exposure to cocaine increased rather than decreased sensitivity to contingency degradation [34, 35]. Overall, besides few exceptions [32, 35, 36], the literature in rodents converges to show that various drugs of abuse shift the balance toward habit.

LIMITATIONS TO THE HABIT THEORY OF ADDICTION

Although drugs of abuse generally promote habit, a very specific set of conditions is typically required to observe habit in rodents. First, the schedule of reinforcement (i.e., random interval) can bias action control toward habit by reducing the contingency and contiguity between response and reinforcement [37–39]. Second, extended operant training can also be required to induce an observable shift toward habit [40–42]. For instance, drug seeking is goal-directed after limited training in the seeking-taking schedule [22–24] but becomes habitual after extended training [24]. Long training is also required to observe the development of alcohol and nicotine habits [11, 20]. Lack of choice seems to be a prerequisite for observing habits during testing. When animals have concurrent access to at

least two rewarded responses, their behavior remains sensitive to outcome devaluation, even after extended training [42–44] or cocaine exposure [34]. Furthermore, the degree of reward predictability seems to play a significant role in habit expression [45–47]. When uncertainty about task contingencies is introduced before testing, this can be sufficient to render habitual behavior, goal-directed again [45, 46]. Finally, expression of habit is typically observed under conditions of extinction. Indeed, when the devalued reinforcer is delivered during reacquisition tests, instrumental responding for drug or nondrug rewards generally becomes sensitive to outcome devaluation [15, 18, 21, 28, 30, 40, 41].

If we consider that behavior remains goal-directed when there is a simple choice between two options, the hypothesis that drug habits contribute to compulsive drug use and ultimately addiction is difficult to reconcile with real-world scenarios, in which drug addicts typically face a multitude of drug and nondrug alternatives [10]. The apparent incompatibility between choice and habit raises another paradox that extends beyond the question of addiction: if this incompatibility were genuine, then how habitual behaviors could be so ubiquitous in everyday life with its rich array of choices and options? In real-world scenarios, habits must somehow be compatible with choice, if only to minimize the costs associated with computationally demanding goal-directed decision-making processes [48, 49]. Another factor limiting the ecological relevance of animal research on habits is that habits have only been observed under extinction conditions, mainly to avoid incentive learning and reengagement of goal-directed control [15, 18, 21, 40, 41]. However, extinction conditions rarely occur in real-world drug use scenarios, in which drug seeking is typically reinforced [10]. Although current animal models appear to fail to demonstrate habit in conditions of higher face validity, the difficulty of observing habit in drug users could also indicate that habit is not an underlying process driving addiction. One way to address this issue is to improve the validity of the habit construct, mainly impeded by the apparent impossibility of observing habit under conditions of choice and reinforcement. However, two recent studies provide new evidence of habit in a drug choice setting and under conditions of reinforcement.

NEW EVIDENCE OF HABITUAL RESPONDING FOR NONDRUG REINFORCERS IN A DRUG CHOICE SETTING

We have recently found that in rats given a choice between a noncaloric solution of saccharin and an intravenous dose of cocaine, responding for saccharin is habitual [50]. Indeed, preference for saccharin was maintained following saccharin devaluation by sensory-specific satiety, in a test conducted under extinction (Fig. 1A, B). In fact, we observed an effect of reward directly reflecting rats' preference for saccharin, but no effect of devaluation on saccharin- and cocaine-seeking behavior (Fig. 1A, B). This insensitivity of saccharin preference to devaluation was replicated using CTA (Fig. 1D, E). Importantly, devaluation of saccharin was verified by showing a reduction of saccharin consumption in the devalued group compared to the non-devalued group for both devaluation methods (Fig. 1C, F).

Another study from our laboratory tested the sensitivity of the rats' preference to changes in the current value of the nondrug option, in conditions of choice and reinforcement [51]. Specifically, water-restricted rats were trained to choose between water and cocaine. Preference was assessed across repeated cycles of water restriction and satiation (Fig. 2A). 1 h or 2 h pre-session access to water (1h- \emptyset and 2h- \emptyset sessions) had no effect on preference and only moderately suppressed water consumption during water trials (Fig. 2A, B). Thus, water was also made available during every intertrial intervals (ITI) of the session (Free-Water condition, FW sessions). This resulted in a drastic suppression of water consumption during water trials, indicating successful devaluation (Fig. 2B). However, rats kept preferentially selecting the water option, even though they consumed little of it. Importantly,

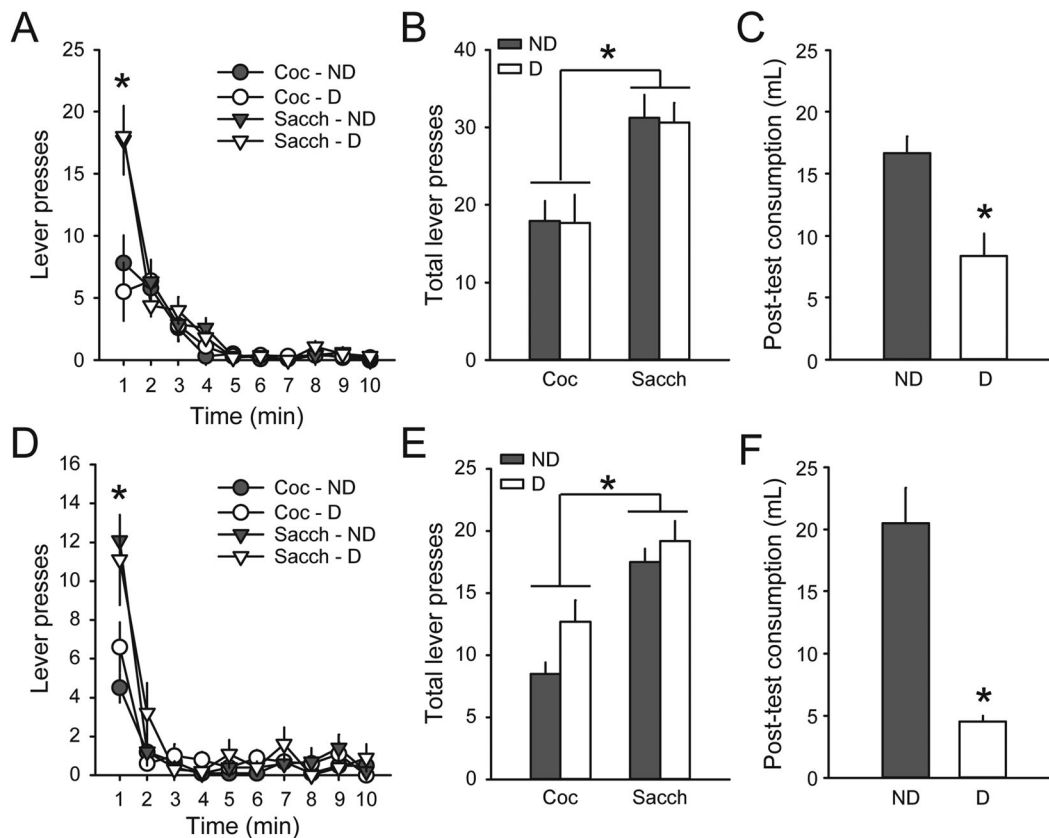


Fig. 1 Habitual preference for saccharin in a drug choice setting. **A–C** Responding for saccharin is not reduced following saccharin devaluation by specific satiety. **A** Rats’ performance on the cocaine and saccharin levers did not differ between the devalued group (D; white) and the non-devalued group (ND; blue) across 1 min time bins in the extinction test. $*p < 0.05$ Coc vs. Sacch. **B** The total number of lever presses was higher on the saccharin lever compared to the cocaine lever but was not affected by devaluation. $*p < 0.05$ Coc vs. Sacch. **C** Saccharin was correctly devalued as measured by a reduction in posttest consumption of saccharin in the D group compared to the ND group. **D–F** Preference for saccharin is also insensitive to saccharin devaluation by CTA. **D, E** Rats responded more on the saccharin lever compared to the cocaine lever but did not differ as a function of devaluation. $*p < 0.05$ Coc vs. Sacch. **F** Devaluation of saccharin was confirmed during the test of consumption immediately after the extinction session. Adapted from [50].

experiencing the devalued outcome during ITI and water trials did not reverse preference toward the still valued drug option by reengaging goal-directed control, indicating that preference for water was habitual and inflexible.

A progressive reversal of preference toward the drug was observed across nine cycles of water restriction and satiation, indicating that preference can only change after repeated training with the novel water value. These results could be well explained in the context of model-based (MB) and model-free (MF) control, used as proxies for goal-directed and habitual control, respectively (Box 2) [48, 52–54]. The slow reversal of preference observed in our study is what would be expected under MF control, which depends on iterative and retrospective learning of an action’s values in a given “state”. Thus, rats may have learned to compute the actions’ value from the start of the session, based on their motivational state. In other words, rats learn to select water when thirsty, and cocaine when sated, without relying on the expected current value of these two rewards. To test this hypothesis, rats were tested again with 1 h water access before the session but not during ITI (1h-Ø; Fig. 2A). Although this condition moderately decreased consumption during water trials, the preference for cocaine increased to 50% and was significantly higher than cocaine preference before devaluation training under the same conditions. These results suggest that during devaluation training, rats learn to use their motivational state as a discriminative cue to predict the most valuable option, under MF control. Alternatively,

since rats became sensitive to the altered outcome value in the presence of an altered interoceptive state (water satiation), it could be argued that rats progressively learned to reengage MB goal-directed control. Yet, rats maintained their preference for water following quinine-induced devaluation, despite a significant suppression of water consumption (Fig. 2A, B), indicating that rats cannot flexibly adjust their preference in response to outcome devaluation using another modality (e.g., taste instead of motivational state). A more parsimonious hypothesis is that rats learned instead to select options according to their motivational state under MF control (i.e., select water when thirsty), without relying on the outcome value per se.

Possible explanations

The results described above are surprising since responding for the nondrug reward was habitual despite choice and reinforcement. In the following subsection, we will discuss possible explanations for these unexpected results.

Both experiments included prior training in the discrete-trial choice schedule to assess preference under baseline conditions. In this procedure, the lever insertion and retraction at each trial constitute salient cues predicting reward availability and delivery, respectively. By reducing uncertainty about reward delivery and alleviating the need for attentional monitoring, these cues can promote the rapid development of habit [47, 55, 56]. Indeed, arbitration between MF and MB control has been suggested to

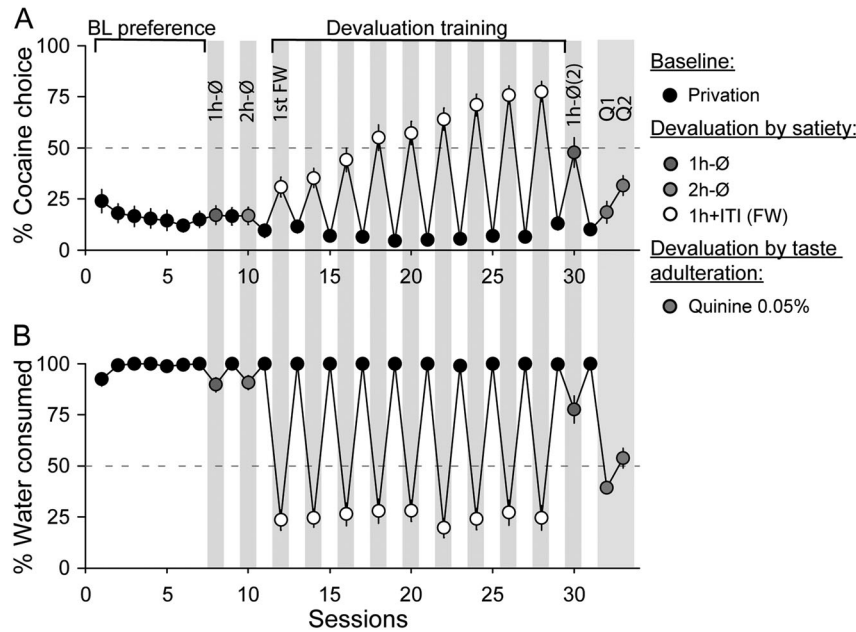
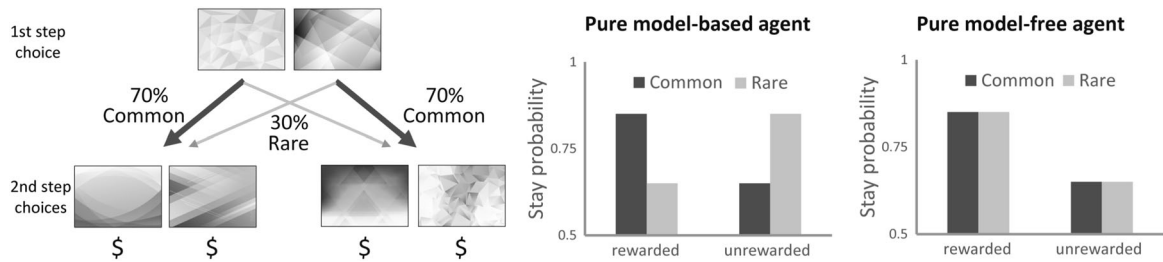


Fig. 2 Inflexible preference for the alternative nondrug reward in a drug choice setting is under habitual, model-free control. Water-restricted rats offered a choice between water and cocaine expressed a robust preference for water (black; baseline preference under water deprivation). Water was then partially devalued with 1 h (1h-Ø, pink) and 2 h free-water access (2h-Ø, purple) before the choice session. Water preference was not affected (A) but there was moderate suppression of water consumption. B Thus, free-water access was also introduced during each intertrial interval (ITI) of choice sessions in addition to the hour of water pre-session access (white; 1 h + ITI, Free-Water FW). Although this condition drastically suppressed water consumption from the first FW session (B), nine sessions were needed to observe a complete reversal of preference (A). Following this devaluation training, 1 h water access was sufficient to raise cocaine preference to 50% in a second 1h-Ø choice session (pink). Finally, devaluation of water by taste adulteration with quinine (blue) only moderately affected preference (A) despite a strong suppression of water consumption (B). Adapted from [51].

Box 2 Model-based and model-free control

Algorithms in reinforcement learning, namely MB and MF learning, have been developed to account for the trade-off between decision speed and accuracy. MB and MF learning formalize the well-documented distinction between goal-directed and habitual behavior, respectively. MB algorithms prospectively learn an internal model of the world, and store a representation of the environment structure (i.e., a cognitive map) in order to compute the expected value of all available courses of actions by iteratively estimating their consequences. MB learning is therefore accurate, but laborious. On the other hand, MF algorithms store and retrieve options “cached values”: the long-run expected value of each action, acquired by iteratively updating actions value through repeated experience of the outcome. This simplified learning model is fast and efficient at the cost of inflexibility: the stored values may be invalid and produce suboptimal choices following changes in task contingencies.

Sequential 2-steps Markov decision task (2-steps task): the 2-steps task teases apart MB and MF control by assessing subjects’ trial-by-trial sensitivity to immediate reward history and task structure (Daw et al. [53]; Gläscher et al. 2010 [130]). In each trial, selecting one stimulus from the 1st-step pair of options results in common and rare transitions to the 2nd-step pairs of options with probabilities of 70% and 30%, respectively. The probabilities of the 2nd-step pairs are reversed following selection of the alternative 1st-step option. 2nd-step options are rewarded according to slowly varying and unpredictable probabilities. Under model-free control, choice does not depend on the transition structure of the task (i.e., common or rare transition) but would only depend on whether the last action was rewarded. In contrast, under MB control, the agent considers both the transition structure of the task (common vs. rare) and prior reward history (rewarded vs. unrewarded). Thus, when the 2nd-step choice is rewarded and follows a rare transition from the 1st-step choice, a MB agent will switch to the alternative 1st-step option (more likely to result in the previously rewarded 2nd-step state) whereas the MF agent will repeat the same 1st-stage choice with no adjustment based on the type of transition.



rely on the relative uncertainty of predictions from each system [52, 57]. In procedures involving discrete trials, the low uncertainty about MF predictions derived from the lever cues through reinforcement learning is hypothesized to favor habit. This could explain why habitual responding for sucrose is observed after only five sessions whereas 8 weeks of training are not sufficient to observe habit when these cues are not available [11, 55]. Therefore, habitual preference in the two studies described above

may be promoted by the structure of the discrete-trial choice procedure. It is noteworthy that studies showing goal-directed choice between two nondrug rewards use self-paced random-ratio or -interval schedules, in absence of reward-predictive cues and thus, under conditions of higher reward uncertainty [34, 42, 44, 58].

The strong initial preference for the alternative nondrug reward in our studies indicates large difference in outcome values [50, 51].

In contrast, studies showing goal-directed choice between two response–outcome associations typically use equally valuable rewards [42–44, 59–62]. In this condition, the brain chooses advantageously by assigning and comparing options value and selecting the response associated with the highest value [63–66]. Consequently, decision-making remains under goal-directed control—driven by a representation of the options' value—when choice outcomes are difficult to distinguish [67]. However, when there is a clear difference in outcome values, choice may not require effortful outcome representation but could instead rely on MF stimulus–response policy, slowly updated based on prior reward history [48]. This is indeed what we observed when assessing rats' preference across repeated cycles of water restriction and satiation [51]. The facilitation of MF control in our experimental choice setting is also in accordance with the arbitration model of Daw et al. based on the relative uncertainty of MB vs. MF predictions [52, 57]. While an increase in task complexity is predicted to favor MB control, the strong difference between value of drug and nondrug rewards combined with the high predictability of reward delivery provided by lever cues should favor MF control.

REFRAMING THE HABIT THEORY OF ADDICTION

In the two studies described above, habitual responding did not promote drug choice but instead favored abstinence. How can we reconcile these results with the habit theory of addiction? In the following section, we will discuss new avenues to reframe the habit theory of addiction by embracing the complexity of (1) drug addicts' decision-making environment and (2) interactions between decision-making processes.

Facing the complexity of drug addicts' environment

The discrete-trial choice procedure developed in our laboratory has been used as a rodent model of addiction to isolate a minority of vulnerable rats that prefer the drug, when the large majority prefers the alternative nondrug reward [68–71]. It is perhaps not surprising that population-wide behavior in rats does not reflect the behavior of the subgroup of individuals losing control over drug use and developing SUD. Future research will assess possible development of habitual cocaine preference in the subset of cocaine-preferring rats.

Although our research departs from the mainstream in showing habitual preference for a nondrug reward in a drug choice setting, there are commonalities with the literature on the role of reward-predictive cues in biasing behavior toward habit. In rodents, it was shown that providing reward-predictive cues—the insertion and retraction of the lever—reduces uncertainty about reward delivery and favors habit [55, 56]. In this context, the lever cue could act as a noncontingent discriminative stimulus signaling the contingency between the response and the reward [72]. Discriminative cues predictive of drug availability have been shown to produce drug seeking in animal models of relapse [72–76]. Interestingly, when smokers are required to choose between cigarette and food rewards, the presentation of discriminative cigarette cues (cigarette pictures) biased preference toward cigarettes, an effect that was not reduced by tobacco devaluation using health warning or satiety [77, 78]. This result suggests that habitual behavior is more strongly bounded by discriminative environmental stimuli and less controlled by the primary drug reinforcement itself.

Noncontingent Pavlovian cues can also directly interact with instrumental reward-seeking behavior, a phenomenon known as “Pavlovian to instrumental transfer” (PIT). Pavlovian cues can elicit a representation of the outcome identity and enhance instrumental responding for that same outcome specifically, independently of the current outcome value (specific-PIT) [42, 79, 80]. Specific-PIT can therefore counteract goal-directed responding by

enhancing responding for an outcome predicted by a cue, despite devaluation of this outcome by satiety [81, 82]. However, the role of PIT in addiction remains unclear [83] and this process is presumably rare in human drug-seeking behavior, which is generally reinforced by contingent drug exposure. Instead, Pavlovian cues are more likely to influence drug-seeking behavior when they are contingent with drug delivery and come to function as conditioned reinforcers (CR), by acquiring motivational salience through repeated pairing with the drug [72, 84]. Although numerous studies demonstrate the fundamental role of CR in producing and maintaining drug-seeking behaviors [72, 75, 85], how resistant habitual behaviors are to changes in CR remains relatively unexplored. More generally, the fundamental role of Pavlovian cues in the control of reward-seeking behaviors remains largely overlooked in tasks employing self-paced free-operant schedules in absence of conditioned and discriminative stimuli.

Because of the multiple interactions between cues, actions and outcomes, task structure plays a fundamental role in the orchestration of associative control during choice behavior. Moving forward, it is fundamental to face the associative complexity underlying drug choice in addiction to understand how interactions between stimuli, actions, and outcomes shape individuals' choices between drug and nondrug rewards.

Facing the complexity of interactions between decision-making processes

The habit theory of addiction is limited by the difficulty of observing habits in real-world settings and evidence that drug-seeking behaviors are primarily goal-directed [5, 10]. It could be argued that behavioral persistence toward a devalued goal results from an excessively strong motivation for the goal rather than from an action executed “out of habit”. Indeed, it was recently suggested that excessive goal-directed control would drive the transition to addiction [10]. Interestingly, evidence suggests that rats showing compulsive-like methamphetamine self-administration (i.e., resistance to footshock punishment) exhibited hyperactivity in the orbitofrontal cortex (OFC) to dorsomedial striatum (DMS) pathways, and lower engagement of the medial prefrontal cortex (mPFC)—ventrolateral striatum circuitry [86]. Furthermore, in a model of optogenetic dopamine neurons self-stimulation [87], it was shown that potentiation of the OFC to dorsal striatum synaptic pathway drives compulsive-like reinforcement [88]. Given the established role of OFC in encoding of value during goal-directed behavior, these results suggest that compulsive-like drug use may be driven by an overestimation of drug value relative to punishment [89]. Furthermore, impairment of executive functioning resulting from drug-induced dysfunctions in PFC activity can disrupt inhibitory control, resulting in an inability to suppress strong motivation after a change in contingencies [89–91]. Together, these studies suggest that compulsive-like drug use is driven by excessive goal-directed motivation for the drug.

Evidence of a shift from ventromedial to dorsolateral striatum in striato-nigro-striatal dopaminergic pathways, which is proposed to underlie the transition from goal-directed to habitual control over drug seeking remains limited. Indeed, studies demonstrating this shift during cocaine self-administration under a second-order schedule of reinforcement did not assess whether behavior was habitual [92]. Although a shift from ventromedial to dorsolateral striatal (DLS) dopamine release has been observed during cocaine self-administration, this shift was suggested to promote refinement of instrumental learning rather than escalated and compulsive-like cocaine seeking [93]. Numerous studies suggest that DMS and DLS are sequentially involved during early and late instrumental training, when behavior is goal-directed or habitual, respectively [94–97]. This dissociation between DMS and DLS has also been reported following ethanol and cocaine self-

administration [11, 24]. Furthermore, dopamine transmission in the DMS and DLS is required for early and late performance of cue-mediated cocaine seeking, respectively [98]. However, the hypothesis of sequential involvement of DMS and DLS across habitual learning has been recently challenged [56] and whether this serial recruitment in dorsostriatal activity is accelerated by drug exposure remains unknown. Clearly, more research is needed to demonstrate a shift in meso-nigro-striatal dopaminergic signaling and dorsostriatal activity in the context of habitual drug-seeking behavior.

Although some neurobiological evidence suggests that addiction is associated with excessive goal-directed drug seeking while other studies seem to indicate a shift toward DLS-dependent drug-seeking habits, drug-related behaviors may not be exclusively habitual or goal directed. There are instances of both goal-directed and habitual behavior in drug addiction. Some strategies developed by drug addicts to acquire money, procure the drug and consume it are undoubtedly goal-directed in that they are highly flexible, driven by expectation of drug effects, and involve careful assessment of risks and benefits [5, 99]. On the other hand, some drug-related behaviors can also be conceived as habitual, for instance, the first cigarette smoked in the morning. Therefore, instead of asking whether drug-seeking behavior is goal-directed or habitual, it may be more relevant to consider exercise of goal-directed control as a gradient and to determine how tilted the balance on that gradient is. However, tasks assessing individual sensitivity to outcome devaluation typically answer a yes-or-no question [100]. In humans, the 2-step task (Box 2) was developed to estimate individual reliance on MB and MF control [48, 52–54] and is more suitable to measure the relative strength of both systems (but see [101]). Using this procedure, several studies have shown correlation between drug use and the strength of MB control [102, 103]. Recent adaptation of this task in rodents [104–106] will provide further information about the relative contribution of MB and MF systems in animal models of addiction [107].

Studies using the 2-step task converge to suggest that goal-directed and habitual control are engaged in parallel and that subjects rely on both systems to make decisions [53, 108]. Several neurocomputational models suggest that habitual and goal-directed processes are intermingled under a hierarchical decision-making structure. Keramati et al. proposed an integrative “plan-until-habit” model in which MF cached values are directly integrated into MB prospective planning [49]. Along the same line, Dezfouli and colleagues proposed that goal-directed choices can be executed under habitual control [109–112]. Alternatively, another model suggests that habitual control can be exerted over goal selection. Selected goals are then reached with deliberation and planning [113]. Although these models propose opposite relationships between goal-directed and habitual systems, all share the assumption that humans constantly and flexibly engage habitual and goal-directed control under hierarchical levels in the decision-making structure. Further blurring the frontier between goal-directed and habitual behaviors, several researchers suggest that habits are by essence goal driven [114, 115].

One key problem of goal-directed, MB strategy is the high computational demand for implementation. In theory, to make decisions under MB control, agents build a decision tree of all possible states and actions and navigate in this “cognitive map” to estimate the long-run worth of each available outcome [48]. In the forest of decision-tree possibilities in real-world settings, considering all the available options is not possible; relevant paths must be somehow preselected [116]. For instance, possible outcomes in a choice situation may be irrelevant and not considered in the first place. We have recently shown in rats that options can be available but not considered in the associative structure of the task, despite the engagement of goal-directed control [117]. In this task, we allowed rats to exert goal-directed

control over the occurrence of choice trials by requiring them to nosepoke in a hole for the presentation of cocaine and saccharin levers (Fig. 3A). As expected, we found that rats preferred saccharin over cocaine but intriguingly, this preference was exclusive in the majority of rats (Fig. 3B). When the interest for saccharin was temporarily lost due to repeated choice (i.e., specific satiety), rats preferred to pause for long periods before reinitiating a choice trial for saccharin, instead of switching to cocaine (Fig. 3C). To explain this suboptimal behavior, we suggested that rats are preferentially associating the initiation of behavioral sequences with saccharin, thereby ignoring the drug reward. These results show that in some situations, choice outcomes can be available but ignored, even when responding is under goal-directed control [117].

These results raise an intriguing question; is it possible to select an option among several choice outcomes without actually choosing between them? Instead of comparing and choosing between options, subjects may only consider the relevant options successively and decide whether to accept or reject them. This is the principle of sequential choice models, which assume that in nature, simultaneous encounters are rare and that mechanisms of choice may be evolutionarily adapted to sequential encounters [118–124]. Applying this model to the discrete-trial choice procedure, choice between drug and nondrug rewards may not involve simultaneous choice with comparison of options value. Instead, only the relevant preferred option would be considered. Since choices are exclusive in this procedure, habitual selection of the nondrug reward with a short latency automatically foregoes the opportunity to select cocaine. Likewise, drug addicts are unlikely to simultaneously choose between drug and nondrug rewards by comparing options values; they may instead decide whether to carry out their drug-seeking sequence. Therefore, experimental settings involving simultaneous choice between options comparable in value in both human and rodent studies may preclude the observation of habit by requiring assessment and comparison of options’ value, thereby reengaging goal-directed control. Yet, this “artificial” choice setting may not represent the true decision-making structure faced by drug users in real-world environment. Although more research is needed to assess the validity of these sequential choice models, this new framework could resolve the challenge of the exponential computational cost of MB strategies in real-world environment and the expression of habit despite choice in our experiments [50, 51], and in the broader context of drug-seeking in addiction.

CONCLUSION

We hope it is clear from this review that habits alone cannot account for the development of compulsive drug use and that drug habits are not necessary [125], nor sufficient [89] to explain the transition to addiction. However, this does not preclude a role for habits in addiction. Then, to what extent are drug habits actually involved? To answer this question, we suffer from several limitations. The structure of our procedures generally favors reengagement of goal-directed control precluding correct assessment of habit. Experiments in animals suffer from a paucity of reward-predictive cues, which does not reflect the sensorial and associative richness of drug addicts’ environment and does not facilitate the development of habit by reducing reinforcement uncertainty. Finally, investigations are limited by too narrow views that drug-seeking behavior should be either habitual or goal-directed. Moving forward, we propose to better design instrumental tasks, in the presence of choice and reward-predictive cues, and under conditions of high reinforcement predictability to favor implementation of simple stimulus–response MF policies. Alternative task structures involving sequential rather than simultaneous choice should also be considered. On a theoretical level, we may need to consider a more complex framework taking

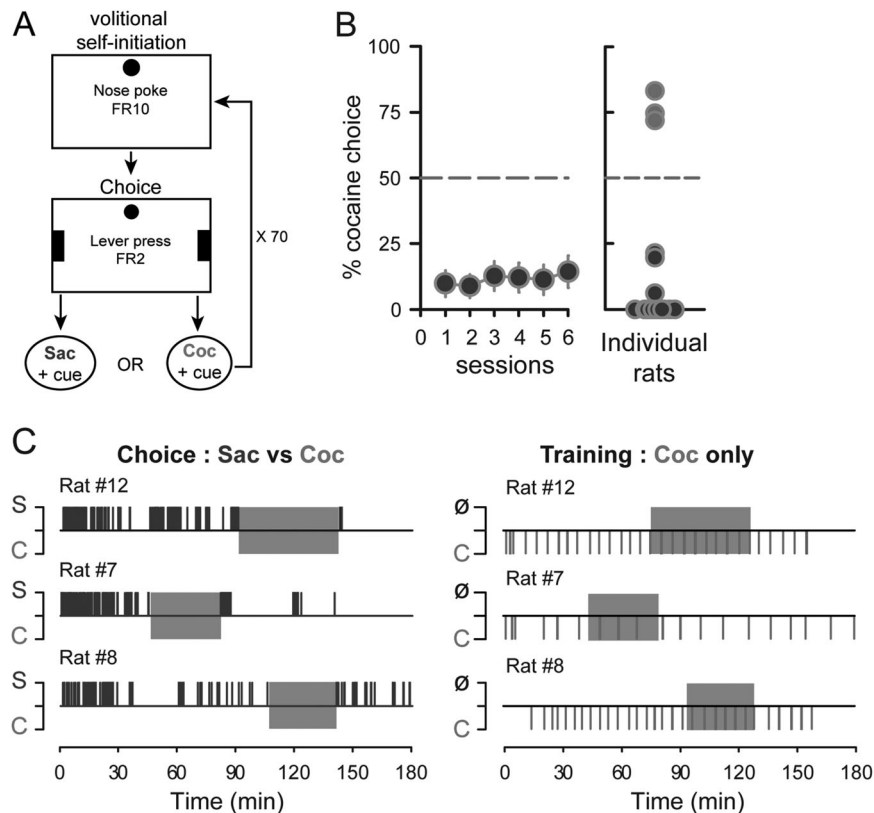


Fig. 3 Rats are oblivious to the cocaine option during self-initiated choice. **A** Rats are required to nosepoke in a hole under a fixed ratio 10 to trigger the presentation of two levers. Two consecutive presses on the left or right lever result in the delivery of saccharin or an intravenous infusion of cocaine, respectively. **B** In this procedure, rats expressed a strong preference for saccharin. Interestingly, this preference was exclusive for a majority of rats (right panel). **C** Analysis of choice patterns reveals that rats choosing saccharin exclusively did so in bouts of varying lengths separated by pauses, during which they did not self-initiated any trial for cocaine, despite transient saccharin devaluation by sensory-specific satiety. This behavior represents an opportunity cost because the duration of pauses is sufficient to earn several cocaine injections (right panel). Adapted from [117].

into account (1) the continuous arbitration between goal-directed and habitual systems, (2) the hierarchical decision-making architectures combining these two systems and (3) alternative sequential decision-making models suggesting that individuals may consider one option at a time when making decisions. Although much remains to be done, our hope is that this review opens up new perspectives to determine the role of habit and choice in addiction.

FUNDING AND DISCLOSURE

This work was supported by the French Research Council (CNRS), the Université de Bordeaux, the French National Agency (ANR-2010-BLAN-1404-01), the Ministère de l'Enseignement Supérieur et de la Recherche (MESR), the Fondation pour la Recherche Médicale (FRM DPA20140629788), and the Peter und Traudl Engelhorn foundation. The authors declare no competing interests.

ACKNOWLEDGEMENTS

We thank Christophe Bernard, Mathieu Louvet, and Eric Wattelet for administrative assistance. We also thank Dr. Patricia Janak for her helpful comments on a previous version of the review, and Emma Chaloux-Pinette for proofreading the paper.

AUTHOR CONTRIBUTIONS

YV drafted the first version of the paper; SHA and YV revised and edited the paper; YV and SHA approved the final version of the paper.

ADDITIONAL INFORMATION

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Redish AD, Jensen S, Johnson A. A unified framework for addiction: vulnerabilities in the decision process. *Behav Brain Sci.* 2008;31:415–87.
2. Everitt BJ, Robbins TW. Drug addiction: updating actions to habits to compulsions ten years on. *Annu Rev Psychol.* 2016;67:23–50.
3. Everitt BJ, Robbins TW. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat Neurosci.* 2005;8:1481–9.
4. Tiffany ST. A cognitive model of drug urges and drug-use behavior: role of automatic and nonautomatic processes. *Psychol Rev.* 1990;97:147–68.
5. Heather N. Is the concept of compulsion useful in the explanation or description of addictive behaviour and experience? *Addict Behav Rep.* 2017;6:15–38.
6. Ostlund SB, Balleine BW. On habits and addiction: an associative analysis of compulsive drug seeking. *Drug Discov Today Dis Model* 2008;5:235–45.
7. Dickinson A, Balleine B. Motivational control of instrumental action. *Anim Learn Behav.* 1994;22:1–18.
8. Balleine BW, Dickinson A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology.* 1998;37:407–19.
9. Hogarth L, Lam-Cassettari C, Pacitti H, Currah T, Mahlberg J, Hartley L, et al. Intact goal-directed control in treatment-seeking drug users indexed by outcome-devaluation and Pavlovian to instrumental transfer: critique of habit theory. *Eur J Neurosci.* 2018;50:2513–2525.
10. Hogarth L. Addiction is driven by excessive goal-directed drug choice under negative affect: translational critique of habit and compulsion theory.

- Neuropsychopharmacology. 2020;45:720–735. <https://doi.org/10.1038/s41386-020-0600-8>.
11. Corbit L, Nie H, Janak P. Habitual alcohol seeking: time course and the contribution of subregions of the dorsal striatum. *Biol Psychiatry*. 2012;72:389–95.
 12. Barker JM, Zhang H, Villafane JJ, Wang TL, Torregrossa MM, Taylor JR. Epigenetic and pharmacological regulation of 5HT3 receptors controls compulsive ethanol seeking in mice. *Eur J Neurosci*. 2014;39:999–1008.
 13. Dickinson A, Wood N, Smith JW. Alcohol seeking by rats: action or habit? *Q J Exp Psychol B* 2002;55:331–48.
 14. Lopez MF, Becker HC, Chandler LJ. Repeated episodes of chronic intermittent ethanol promote insensitivity to devaluation of the reinforcing effect of ethanol. *Alcohol*. 2014;48:639–45.
 15. Mangieri RA, Cofresi RU, Gonzales RA. Ethanol seeking by long evans rats is not always a goal-directed behavior. *PLoS ONE*. 2012;7:1–13.
 16. Mangieri RA, Cofresi RU, Gonzales RA. Ethanol exposure interacts with training conditions to influence behavioral adaptation to a negative instrumental contingency. *Front Behav Neurosci*. 2014;8:220.
 17. Corbit LH, Nie H, Janak PH. Habitual responding for alcohol depends upon both AMPA and D2 receptor signaling in the dorsolateral striatum. *Front Behav Neurosci*. 2014;8:1–9.
 18. Miles FJ, Everitt BJ, Dickinson A. Oral cocaine seeking by rats: action or habit? *Behav Neurosci*. 2003;117:927–38.
 19. Leong KC, Berini CR, Ghee SM, Reichel CM. Extended cocaine-seeking produces a shift from goal-directed to habitual responding in rats. *Physiol Behav*. 2016;164:330–5.
 20. Clemens KJ, Lay BP, Holmes NM. Extended nicotine self-administration increases sensitivity to nicotine, motivation to seek nicotine and the reinforcing properties of nicotine-paired cues. *Addict Biol*. 2015;22:400–410.
 21. Loughlin A, Funk D, Coen K, Lê AD. Habitual nicotine-seeking in rats following limited training. *Psychopharmacology*. 2017;234:2619–2629.
 22. Olmstead MC, Lafonda MV, Everitt BJ, Dickinson A. Cocaine seeking by rats is a goal-directed action. *Behav Neurosci*. 2001;115:394–402.
 23. Hutcheson DM, Everitt BJ, Robbins TW, Dickinson A. The role of withdrawal in heroin addiction: enhances reward or promotes avoidance? *Nat Neurosci*. 2001;4:943–7.
 24. Zapata A, Minney VL, Shippenberg TS. Shift from goal-directed to habitual cocaine seeking after prolonged experience in rats. *J Neurosci*. 2010;30:15457–63.
 25. Renteria R, Baltz ET, Gremel CM. Chronic alcohol exposure disrupts top-down control over basal ganglia action selection to produce habits. *Nat Commun*. 2018;9:1–11.
 26. Corbit LH, Chieng BC, Balleine BW. Effects of repeated cocaine exposure on habit learning and reversal by N-acetylcysteine. *Neuropsychopharmacology*. 2014;39:1893–901.
 27. LeBlanc KH, Maidment NT, Ostlund SB. Repeated cocaine exposure facilitates the expression of incentive motivation and induces habitual control in rats. *PLoS ONE*. 2013;8:e61355:1–10.
 28. Nelson A, Killcross S. Amphetamine exposure enhances habit formation. *J Neurosci*. 2006;26:3805–12.
 29. Nordquist RE, Voorn P, de Mooij-van Malsen JG, Joosten RNJMA, Pennartz CMA, Vanderschuren LJMJ. Augmented reinforcer value and accelerated habit formation after repeated amphetamine treatment. *Eur Neuropsychopharmacol*. 2007;17:532–40.
 30. Nelson AJD, Killcross S, LeBlanc KH. Accelerated habit formation following amphetamine exposure is reversed by D 1, but enhanced by D 2, receptor antagonists. *Front Neurosci*. 2013;7:1–13.
 31. Schmitzer-Torbert N, Apostolidis S, Amoa R, O'Rear C, Kaster M, Stowers J, et al. Post-training cocaine administration facilitates habit learning and requires the infralimbic cortex and dorsolateral striatum. *Neurobiol Learn Mem*. 2015;118:105–12.
 32. Shiflett MW. The effects of amphetamine exposure on outcome-selective Pavlovian-instrumental transfer in rats. *Psychopharmacology*. 2012;223:361–70.
 33. Gourley SL, Olevska A, Gordon J, Taylor JR. Cytoskeletal determinants of stimulus-response habits. *J Neurosci*. 2013;33:11811–6.
 34. Halbout B, Liu AT, Ostlund SB. A closer look at the effects of repeated cocaine exposure on adaptive decision-making under conditions that promote goal-directed control. *Front Psychiatry*. 2016;7:1–12.
 35. Phillips GD, Vugler A. Effects of sensitization on the detection of an instrumental contingency. *Pharmacol Biochem Behav*. 2011;100:48–58.
 36. Son JH, Latimer C, Keefe KA. Impaired formation of stimulus-response, but not action-outcome, associations in rats with methamphetamine-induced neurotoxicity. *Neuropsychopharmacology*. 2011;36:2441–51.
 37. Dickinson A, Nicholas DJ, Adams CD. The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Q J Exp Psychol Sect B*. 1983;35:35–51.
 38. Derusso AL, Fan D, Gupta J, Shelest O, Costa RM, Yin HH. Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. *Front Integr Neurosci*. 2010;4:1–8.
 39. Urcelay GP, Jonkman S. Delayed rewards facilitate habit formation delayed rewards facilitate habit formation. *J Exp Psychol Anim Learn Cogn*. 2019;45:413–421.
 40. Adams CD. Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q J Exp Psychol Sect B*. 1982;34:77–98.
 41. Dickinson A. Actions and habits: the development of behavioural autonomy. *Philos Trans R Soc B Biol Sci*. 1985;308:67–78.
 42. Holland PC. Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *J Exp Psychol Anim Behav Process*. 2004;30:104–17.
 43. Colwill RM, Triola SM. Instrumental responding remains under the control of the consequent outcome after extended training. *Behav Process*. 2002;57:51–64.
 44. Kosaki Y, Dickinson A. Choice and contingency in the development of behavioral autonomy during instrumental conditioning. *J Exp Psychol Anim Behav Process*. 2010;36:334–42.
 45. Trask S, Shipman ML, Green JT, Bouton ME. Some factors that restore goal-direction to a habitual behavior. *Neurobiol Learn Mem*. 2020;169:107161.
 46. Bouton ME, Broomer MC, Rey CN, Thraillkill EA. Unexpected food outcomes can return a habit to goal-directed action. *Neurobiol Learn Mem*. 2020;169:1–9.
 47. Thraillkill EA, Trask S, Vidal P, Alcalá JA, Bouton ME. Stimulus control of actions and habits: a role for reinforcer predictability and attention in the development of habitual behavior. *J Exp Psychol Anim Learn Cogn*. 2018;44:370–84.
 48. Dolan RJ, Dayan P. Goals and habits in the brain. *Neuron*. 2013;80:312–25.
 49. Keramati M, Smittenaar P, Dolan RJ, Dayan P. Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proc Natl Acad Sci USA*. 2016;113:12868–73.
 50. Vandaele Y, Guillem K, Ahmed SH, Ahmed SH. Habitual preference for the nondrug reward in a drug choice setting. *Front Behav Neurosci*. 2020;14:1–9.
 51. Vandaele Y, Vouillac-Mendoza C, Ahmed SH. Inflexible habitual decision-making during choice between cocaine and a nondrug alternative. *Transl Psychiatry*. 2019;9:109.
 52. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci*. 2005;8:1704–11.
 53. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. *Neuron*. 2011;69:1204–15.
 54. Doya K, Samejima K, Katagiri K, Kawato M. Multiple model-based reinforcement learning. *Neural Comput*. 2002;14:1347–69.
 55. Vandaele Y, Pribut HJ, Janak PH. Lever insertion as a salient stimulus promoting insensitivity to outcome devaluation. *Front Integr Neurosci*. 2017;11:1–13.
 56. Vandaele Y, Mahajan NR, Ottenheimer DJ, Richard JM, Mysore SP, Janak PH. Distinct recruitment of dorsomedial and dorsolateral striatum erodes with extended training. *Elife*. 2019;8:1–29.
 57. Lee SW, Shimojo S, O'Doherty JP. Neural computations underlying arbitration between model-based and model-free learning. *Neuron*. 2014;81:687–99.
 58. Colwill RM, Rescorla RA. Instrumental responding remains sensitive to reinforcer devaluation after extensive training. *J Exp Psychol Anim Behav Process*. 1985;11:520–36.
 59. Parkes SL, Balleine BW. Incentive memory: evidence the basolateral amygdala encodes and the insular cortex retrieves outcome values to guide choice between goal-directed actions. *J Neurosci*. 2013;33:8753–63.
 60. Parkes SL, Bradfield LA, Balleine BW. Interaction of insular cortex and ventral striatum mediates the effect of incentive memory on choice between goal-directed actions. *J Neurosci*. 2015;35:6464–71.
 61. Balleine BW, Killcross AS, Dickinson A. The effect of lesions of the basolateral amygdala on instrumental conditioning. *J Neurosci*. 2003;23:666–75.
 62. Corbit LH, Balleine BW. The role of prefrontal cortex in instrumental conditioning. *Behav Brain Res*. 2003;146:145–57.
 63. Glimcher PW, Rustichini A. Neuroeconomics: the consilience of brain and decision. *Science*. 2004;306:447–52.
 64. Rangel A, Camerer C, Montague PR. A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci*. 2008;9:545–56.
 65. Rushworth MF, Mars RB, Summerfield C. General mechanisms for making decisions? *Curr Opin Neurobiol*. 2009;19:75–83.
 66. Rangel A, Hare T. Neural computations associated with goal-directed choice. *Curr Opin Neurobiol*. 2010;20:262–70.
 67. Keramati M, Dezfouli A, Piray P. Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Comput Biol*. 2011;7:e1002055.
 68. Lenoir M, Augier E, Vouillac C, Ahmed SH. A choice-based screening method for compulsive drug users in rats. *Curr Protoc Neurosci*. 2013;1:1–17.
 69. Cantin L, Lenoir M, Augier E, Vanhille N, Dubreucq S, Serre F, et al. Cocaine is low on the value ladder of rats: Possible evidence for resilience to addiction. *PLoS ONE*. 2010;5:e11592:1–14.

70. Lenoir M, Serre F, Cantin L, Ahmed SH. Intense sweetness surpasses cocaine reward. *PLoS ONE*. 2007;2:e698:1–10.
71. Ahmed SH. The science of making drug-addicted animals. *Neuroscience*. 2012;211:107–25.
72. Namba MD, Tomek SE, Olive MF, Beckmann JS, Gipson CD. The winding road to relapse: forging a new understanding of cue-induced reinstatement models and their associated neural mechanisms. *Front Behav Neurosci*. 2018;12:1–22.
73. Weiss F, Maldonado-Vlaar CS, Parsons LH, Kerr TM, Smith DL, Ben-Shahar O. Control of cocaine-seeking behavior by drug-associated stimuli in rats: Effects on recovery of extinguished operant-responding and extracellular dopamine levels in amygdala and nucleus accumbens. *Proc Natl Acad Sci USA*. 2000;97:4321–6.
74. Weiss F, Martin-Fardon R, Ciccocioppo R, Kerr TM, Smith DL, Ben-Shahar O. Enduring resistance to extinction of cocaine-seeking behavior induced by drug-related cues. *Neuropsychopharmacology*. 2001;25:361–72.
75. Shaham Y, Shalev U, Lu L, de Wit H, Stewart J. The reinstatement model of drug relapse: history, methodology and major findings. *Psychopharmacology*. 2003;168:3–20.
76. Crombag HS, Bossert JM, Koya E, Shaham Y. Context-induced relapse to drug seeking: a review. *Philos Trans R Soc B Biol Sci*. 2008;363:3233–43.
77. Hogarth L, Chase HW. Parallel goal-directed and habitual control of human drug-seeking: implications for dependence vulnerability. *J Exp Psychol Anim Behav Process*. 2011;37:261–76.
78. Hogarth L. Goal-directed and transfer-cue-elicited drug-seeking are dissociated by pharmacotherapy: evidence for independent additive controllers. *J Exp Psychol Anim Behav Process*. 2012;38:266–78.
79. Corbit LH, Janak PH, Balleine BW. General and outcome-specific forms of Pavlovian-instrumental transfer: the effect of shifts in motivational state and inactivation of the ventral tegmental area. *Eur J Neurosci*. 2007;26:3141–9.
80. Rescorla RA. Transfer of instrumental control mediated by a devalued outcome. *Anim Learn Behav*. 1994;22:27–33.
81. Watson P, Wiers RW, Hommel B, de Wit S. Working for food you don't desire. Cues interfere with goal-directed food-seeking. *Appetite* 2014;79:139–48.
82. Van Steenbergen H, Watson P, Wiers RW, Hommel B, de Wit S. Dissociable corticostriatal circuits underlie goal-directed vs. cue-elicited habitual food seeking after satiation: evidence from a multimodal MRI study. *Eur J Neurosci*. 2017;46:1815–1827.
83. Lamb RJ, Schindler W, Pinkston JW. Conditioned stimuli's role in relapse: pre-clinical research on pavlovian instrumental transfer. *Psychopharmacology*. 2016;233:1933–44.
84. Kelleher RT, Gollub LR. A review of positive conditioned reinforcement. *J Exp Anal Behav*. 1962;5:543–97.
85. Bossert JM, Marchant NJ, Calu DJ, Shaham Y. The reinstatement model of drug relapse: recent neurobiological findings, emerging research topics, and translational research. *Psychopharmacology*. 2013;229:453–76.
86. Hu Y, Salmeron BJ, Krasnova IN, Gu H, Lu H, Bonci A, et al. Compulsive drug use is associated with imbalance of orbitofrontal- And prelimbic-striatal circuits in punishment-resistant individuals. *Proc Natl Acad Sci USA*. 2019;116:9066–71.
87. Pascoli V, Terrier J, Hiver A, Lüscher C. Sufficiency of mesolimbic dopamine neuron stimulation for the progression to addiction. *Neuron*. 2015;88:1054–66.
88. Pascoli V, Hiver A, Van Zessen R, Loureiro M, Achargui R, Harada M, et al. Stochastic synaptic plasticity underlying compulsion in a model of addiction. *Nature*. 2018;564:366–71.
89. Lüscher C, Robbins TW, Everitt BJ. The transition to compulsion in addiction. *Nat Rev Neurosci*. 2020;21:247–63.
90. Goldstein RZ, Volkow ND. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat Rev Neurosci*. 2012;12:652–69.
91. Baler RD, Volkow ND. Drug addiction: the neurobiology of disrupted self-control. *Trends Mol Med*. 2006;12:559–66.
92. Belin D, Everitt BJ. Cocaine seeking habits depend upon dopamine-dependent serial connectivity linking the ventral with the dorsal striatum. *Neuron*. 2008;57:432–41.
93. Willuhn I, Burgeno LM, Everitt BJ, Phillips PEM. Hierarchical recruitment of phasic dopamine signaling in the striatum during the progression of cocaine use. *Proc Natl Acad Sci USA*. 2012;109:20703–8.
94. Yin HH, Ostlund SB, Knowlton BJ, Balleine BW. The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci*. 2005;22:513–23.
95. Yin HH, Knowlton BJ, Balleine BW. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci*. 2004;19:181–9.
96. Yin HH, Knowlton BJ. The role of the basal ganglia in habit formation. *Nat Rev Neurosci*. 2006;7:464–76.
97. Balleine BW, Liljeholm M, Ostlund SB. The integrative function of the basal ganglia in instrumental conditioning. *Behav Brain Res*. 2009;199:43–52.
98. Murray JE, Belin D, Everitt BJ. Double dissociation of the dorsomedial and dorsolateral striatal control over the acquisition and performance of cocaine seeking. *Neuropsychopharmacology*. 2012;37:2456–66.
99. Preble E. Taking care of business—the heroin user's life on the street. *Int J Addict*. 1969;4:1–24.
100. Schreiner DC, Renteria R, Gremel CM. Fractionating the all-or-nothing definition of goal-directed and habitual decision-making. *J Neurosci Res*. 2019;98:998–1006.
101. Feher da Silva C, Hare TA. Humans primarily use model-based inference in the two-stage task. *Nat Hum Behav*. 2020. <https://doi.org/10.1038/s41562-020-0905-y>.
102. Byrne KA, Otto AR, Pang B, Patrick CJ, Worthy DA. Substance use is associated with reduced devaluation sensitivity. *Cogn Affect Behav Neurosci*. 2019;19:40–55.
103. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *Elife*. 2016;5:e11305:1–24.
104. Miller KJ, Botvinick MM, Brody CD. Dorsal hippocampus contributes to model-based planning. *Nat Neurosci*. 2017;20:1269–76.
105. Groman SM, Massi B, Mathias SR, Lee D, Taylor JR. Model-free and model-based influences in addiction-related behaviors. *Biol Psychiatry*. 2019;85:936–45.
106. Groman SM, Massi B, Mathias SR, Curry DW, Lee D, Taylor JR. Neurochemical and behavioral dissections of decision-making in a rodent multistage task. *J Neurosci*. 2019;39:295–306.
107. Fraser KM, Janak PH. How does drug use shift the balance between model-based and model-free control of decision making? *Biol Psychiatry*. 2019;85:886–8.
108. Otto AR, Gershman SJ, Markman AB, Daw ND. The curse of planning. *Psychol Sci*. 2013;24:751–61.
109. Dezfouli A, Balleine BW. Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Comput Biol*. 2013;9:e1003364.
110. Dezfouli A, Balleine BW. Habits, action sequences and reinforcement learning. *Eur J Neurosci*. 2012;35:1036–51. <https://doi.org/10.1038/s41386-020-0600-8>.
111. Dezfouli A, Lingawi NW, Balleine BW. Habits as action sequences: hierarchical action control and changes in outcome value. *Philos Trans R Soc L B Biol Sci*. 2014;369:20130482.
112. Balleine BW, Dezfouli A. Hierarchical action control: adaptive collaboration between actions and habits. *Front Psychol*. 2019;10:1–13.
113. Cushman F, Morris A. Habitual control of goal selection in humans. *Proc Natl Acad Sci USA*. 2015;112:13817–22.
114. Wood W, Neal DT. A new look at habits and the habit—goal interface. *Psychol Rev*. 2007;114:843–63.
115. Kruglanski AW, Szumowska E. Habitual behavior is goal-driven. *Perspect Psychol Sci*. 2020. <https://doi.org/10.1177/1745691620917676>.
116. Huys QJM, Eshel N, O'Nions E, Sheridan L, Dayan P, Roiser JP. Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput Biol*. 2012;8:1–13.
117. Vandaale Y, Vouillac-Mendoza C, Ahmed SH. Cocaine falls into oblivion during volitional initiation of choice trials. *Addict Biol*. 2020. In press.
118. Shapiro MS, Siller S, Kacelnik A. Simultaneous and sequential choice as a function of reward delay and magnitude: normative, descriptive and process-based models tested in the European Starling (*Sturnus vulgaris*). *J Exp Psychol Anim Behav Process*. 2008;34:75–93.
119. Freidin E, Aw J, Kacelnik A. Sequential and simultaneous choices: testing the diet selection and sequential choice models. *Behav Process*. 2009;80:218–23.
120. Freidin E, Kacelnik A. Rational choice, context dependence, and the value of information in European starlings (*Sturnus vulgaris*). *Science*. 2011;334:1000–2.
121. Vasconcelos M, Monteiro T, Aw J, Kacelnik A. Choice in multi-alternative environments: a trial-by-trial implementation of the sequential choice model. *Behav Process*. 2010;84:435–9.
122. Vasconcelos M, Monteiro T, Kacelnik A. Context-dependent preferences in starlings: linking ecology, foraging and choice. *PLoS ONE*. 2013;8:1–8.
123. Mobbs D, Trimmer PC, Blumstein DT, Dayan P. Foraging for foundations in decision neuroscience: insights from ethology. *Nat Rev Neurosci*. 2018;19:419–27.
124. Grace RC. Acquisition of choice in concurrent chains: assessing the cumulative decision model. *Behav Process*. 2016;126:82–93.
125. Singer BF, Fadanelli M, Kawa AB, Robinson TE. Are cocaine-seeking "habits" necessary for the development of addiction-like behavior in rats? *J Neurosci*. 2017;38:60–73.
126. Colwill RM. An associative analysis of instrumental learning. *Curr Dir Psychol Sci*. 1993;2:111–6.

127. Dickinson A, Mulatero CW. Reinforcer specificity of the suppression of instrumental performance on a non-contingent schedule. *Behav Processes*. 1989;19:167–80.
128. Rescorla RA. A Pavlovian analysis of goal-directed behavior. *American Psychologist* 1987;42:119–29.
129. Adams CD, Dickinson A. Instrumental responding following reinforcer devaluation. *Q J Exp Psychol Sect B Comp Physiol Psychol*. 1981;33:109–121.
130. Gläscher J, Daw N, Dayan P, O'Doherty JP. States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning. *Neuron* 2010;66:585–95.