



HAL
open science

Habit, choice and addiction

Y. Vandaele, Serge Ahmed H.

► **To cite this version:**

Y. Vandaele, Serge Ahmed H.. Habit, choice and addiction. Neuropsychopharmacology, In press, 10.1038/s41386-020-00899-y . hal-03005465

HAL Id: hal-03005465

<https://hal.science/hal-03005465>

Submitted on 25 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Habit, Choice and Addiction

Vandaele Y.^{1*} and Ahmed S.H.^{2,3}

¹Lausanne University Hospital, Department of Psychiatry, 1008 Prilly, Switzerland ; ²Université de Bordeaux, Institut des Maladies Neurodégénératives, UMR 5293, 146 rue Léo-Saignat, F-33000 Bordeaux, France ; ³CNRS, Institut des Maladies Neurodégénératives, UMR 5293, 146 rue Léo-Saignat, F-33000 Bordeaux, France.

*corresponding author: Youna Vandaele, PhD,
Lab on the Neurobiology of Addictive and Eating Disorders,
CHUV-Lausanne University, Department of Psychiatry,
Switzerland

Email : youna.vandaele@chuv.ch

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

1 **Introduction**

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

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

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

34 **Drugs promote habit**

35 A large number of studies in rodents show that drugs of abuse promote habit. Following drug self-
36 administration training, drugs can be devalued using either sensory-specific satiety or conditioned
37 taste aversion (CTA) before responding for the drug is tested under extinction (box 1). Using this
38 procedure, it was shown that responding for ethanol [11–17], cocaine [18,19] and nicotine [20,21]
39 becomes habitual after various length of training. In some studies, the transition to habit was faster
40 for the drug compared to a non-drug reward suggesting stronger facilitation of habit formation for
41 drug-seeking [11,13,15,18,21]. Interestingly, studies in which rats are trained to self-administer
42 cocaine or heroin in a seeking-taking schedule (e.g. heterogeneous chains; seeking RI30 – taking FR1
43 on separate levers) reveal that rats correctly encode the contingency between the seeking response,

1 the taking response and the outcome, indicating that their behavior is under goal-directed control
2 [22,23]. However, it was also shown that the cocaine seeking response becomes insensitive to
3 extinction of the cocaine taking response following extended self-administration training, suggesting
4 a shift to habitual control [24].

5
6 Numerous studies show that passive drug exposure is sufficient to promote habitual responding for
7 non-drug rewards. For instance, while lever pressing for a solution of 20% sucrose remains under
8 goal-directed control after 8 weeks of training, home-cage access to ethanol during instrumental
9 training renders the behavior habitual [11]. Ethanol-induced facilitation of habitual responding for
10 food was also found following chronic intermittent exposure to ethanol vapor [25]. Passive cocaine
11 [26,27] or amphetamine [28–30] exposure also rendered responding for a non-drug reward
12 insensitive to devaluation by specific satiety or CTA. Interestingly, even limited post-training
13 exposure to cocaine was sufficient to observe habitual responding for food rewards [31], a results
14 not replicated with amphetamine [32]. Drug-induced facilitation of habit was also demonstrated in
15 studies showing insensitivity to degradation of instrumental contingency (box 1) following ethanol
16 exposure [16] or repeated injections of cocaine [33]. However, two studies have found that exposure
17 to cocaine increased rather than decreased sensitivity to contingency degradation [34,35]. Overall,
18 besides few exceptions [32,35,36], the literature in rodents converges to show that various drugs of
19 abuse shift the balance toward habit.

20 **Limitations to the habit theory of addiction**

21 Although drugs of abuse generally promote habit, a very specific set of conditions is typically
22 required to observe habit in rodents. First, the schedule of reinforcement (i.e. random interval) can
23 bias action control toward habit by reducing the contingency and contiguity between response and
24 reinforcement [37–39]. Second, extended operant training can also be required to induce an
25 observable shift toward habit [40–42]. For instance, drug seeking is goal-directed after limited
26 training in the seeking-taking schedule [22–24] but becomes habitual after extended training [24].
27 Long training is also required to observe the development of alcohol and nicotine habits [11,20]. Lack
28 of choice seems to be a prerequisite for observing habits during testing. When animals have
29 concurrent access to at least two rewarded responses, their behavior remains sensitive to outcome
30 devaluation, even after extended training [42–44] or cocaine exposure [34]. Furthermore, the degree
31 of reward predictability seems to play a significant role in habit expression [45–47]. When
32 uncertainty about task contingencies is introduced before testing, this can be sufficient to render
33 habitual behavior, goal-directed again [45,46]. Finally, expression of habit is typically observed under
34 conditions of extinction. Indeed, when the devalued reinforcer is delivered during reacquisition tests,
35 instrumental responding for drug or non-drug rewards generally becomes sensitive to outcome
36 devaluation [15,18,21,28,30,40,41].

37 If we consider that behavior remains goal-directed when there is a simple choice between two
38 options, the hypothesis that drug habits contribute to compulsive drug use and ultimately addiction
39 is difficult to reconcile with real-world scenarios, in which drug addicts typically face a multitude of
40 drug and non-drug alternatives [10]. The apparent incompatibility between choice and habit raises
41 another paradox that extends beyond the question of addiction: if this incompatibility were genuine,
42 then how habitual behaviors could be so ubiquitous in everyday life with its rich array of choices and
43 options? In real-world scenarios, habits must somehow be compatible with choice, if only to
44 minimize the costs associated with computationally demanding goal-directed decision-making

1 processes [48,49]. Another factor limiting the ecological relevance of animal research on habits is
2 that habits have only been observed under extinction conditions, mainly to avoid incentive learning
3 and re-engagement of goal-directed control [15,18,21,40,41]. However, extinction conditions rarely
4 occur in real-world drug use scenarios, in which drug-seeking is typically reinforced [10]. Although
5 current animal models appear to fail to demonstrate habit in conditions of higher face-validity, the
6 difficulty of observing habit in drug users could also indicate that habit is not an underlying process
7 driving addiction. One way to address this issue is to improve the validity of the habit construct,
8 mainly impeded by the apparent impossibility of observing habit under conditions of choice and
9 reinforcement. However, two recent studies provide new evidence of habit in a drug choice setting
10 and under conditions of reinforcement.

11 New evidence of habitual responding for non-drug reinforcers in a drug choice 12 setting

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

22 Another study from our laboratory tested the sensitivity of the rats' preference to changes in the
23 current value of the non-drug option, in conditions of choice and reinforcement [51]. Specifically,
24 water-restricted rats were trained to choose between water and cocaine. Preference was assessed
25 across repeated cycles of water restriction and satiation (Figure 2A). 1h or 2h pre-session access to
26 water (1h- \emptyset and 2h- \emptyset sessions) had no effect on preference and only moderately suppressed water
27 consumption during water trials (Figure 2A-B). Thus, water was also made available during every
28 inter-trial intervals (ITI) of the session (Free Water condition, FW sessions). This resulted in a drastic
29 suppression of water consumption during water trials, indicating successful devaluation (Figure 2B).
30 However, rats kept preferentially selecting the water option, even though they consumed little of it.
31 Importantly, experiencing the devalued outcome during ITI and water trials did not reverse
32 preference toward the still valued drug option by re-engaging goal-directed control, indicating that
33 preference for water was habitual and inflexible.

34 A progressive reversal of preference toward the drug was observed across 9 cycles of water
35 restriction and satiation, indicating that preference can only change after repeated training with the
36 novel water value. These results could be well explained in the context of model-based and model-
37 free control, used as proxies for goal-directed and habitual control, respectively (box 2) [48,52–54].
38 The slow reversal of preference observed in our study is what would be expected under model-free
39 control, which depends on iterative and retrospective learning of an action's values in a given 'state'.
40 Thus, rats may have learned to compute the actions' value from the start of the session, based on
41 their motivational state. In other words, rats learn to select water when thirsty, and cocaine when
42 sated, without relying on the expected current value of these two rewards. To test this hypothesis,
43 rats were tested again with 1h water access before the session but not during ITI (1h- \emptyset ; Figure 2A).

1 Although this condition moderately decreased consumption during water trials, the preference for
2 cocaine increased to 50% and was significantly higher than cocaine preference before devaluation
3 training under the same conditions. These results suggest that during devaluation training, rats learn
4 to use their motivational state as a discriminative cue to predict the most valuable option, under
5 model-free control. Alternatively, since rats became sensitive to the altered outcome value in the
6 presence of an altered interoceptive state (water satiation), it could be argued that rats progressively
7 learned to reengage model-based goal-directed control. Yet, rats maintained their preference for
8 water following quinine-induced devaluation, despite a significant suppression of water consumption
9 (Figure 2A-B), indicating that rats cannot flexibly adjust their preference in response to outcome
10 devaluation using another modality (e.g. taste instead of motivational state). A more parsimonious
11 hypothesis is that rats learned instead to select options according to their motivational state under
12 model-free control (i.e. select water when thirsty), without relying on the outcome value *per se*.

13 **a. Possible explanations**

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

17 Both experiments included prior training in the discrete-trial choice schedule to assess preference
18 under baseline conditions. In this procedure, the lever insertion and retraction at each trial constitute
19 salient cues predicting reward availability and delivery, respectively. By reducing uncertainty about
20 reward delivery and alleviating the need for attentional monitoring, these cues can promote the
21 rapid development of habit [47,55,56]. Indeed, arbitration between model-free and model-based
22 control has been suggested to rely on the relative uncertainty of predictions from each system
23 [52,57]. In procedures involving discrete trials, the low uncertainty about model-free predictions
24 derived from the lever cues through reinforcement learning is hypothesized to favor habit. This could
25 explain why habitual responding for sucrose is observed after only 5 sessions whereas 8 weeks of
26 training are not sufficient to observe habit when these cues are not available [11,55]. Therefore,
27 habitual preference in the two studies described above may be promoted by the structure of the
28 discrete-trial choice procedure. It is noteworthy that studies showing goal-directed choice between
29 two non-drug rewards use self-paced random-ratio or -interval schedules, in absence of reward-
30 predictive cues and thus, under conditions of higher reward uncertainty [34,42,44,58].

31 The strong initial preference for the alternative non-drug reward in our studies indicates large
32 difference in outcome values [50,51]. In contrast, studies showing goal-directed choice between two
33 Response-Outcome associations typically use equally valuable rewards [42–44,59–62]. In this
34 condition, the brain chooses advantageously by assigning and comparing options value and selecting
35 the response associated with the highest value [63–66]. Consequently, decision-making remains
36 under goal-directed control – driven by a representation of the options’ value – when choice
37 outcomes are difficult to distinguish [67]. However, when there is a clear difference in outcome
38 values, choice may not require effortful outcome representation but could instead rely on model-
39 free stimulus-response policy, slowly updated based on prior reward history [48]. This is indeed what
40 we observed when assessing rats’ preference across repeated cycles of water restriction and
41 satiation [51]. The facilitation of model-free control in our experimental choice setting is also in
42 accordance with the arbitration model of Daw et al. (2005), based on the relative uncertainty of
43 model-based vs model-free predictions [52,57]. While an increase in task complexity is predicted to

1 favor model-based control, the strong difference between value of drug and non-drug rewards
2 combined with the high predictability of reward delivery provided by lever cues should favor model-
3 free control.

4 **Reframing the habit theory of addiction**

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

10 **a. Facing the complexity of drug addicts' environment**

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

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

30 Non-contingent Pavlovian cues can also directly interact with instrumental reward-seeking behavior,
31 a phenomenon known as “Pavlovian to instrumental transfer” (PIT). Pavlovian cues can elicit a
32 representation of the outcome identity and enhance instrumental responding for that same outcome
33 specifically, independently of the current outcome value (specific-PIT) [42,79,80]. Specific-PIT can
34 therefore counteract goal-directed responding by enhancing responding for an outcome predicted by
35 a cue, despite devaluation of this outcome by satiety [81,82]. However, the role of PIT in addiction
36 remains unclear [83] and this process is presumably rare in human drug seeking behavior, which is
37 generally reinforced by contingent drug exposure. Instead, Pavlovian cues are more likely to
38 influence drug-seeking behavior when they are contingent with drug delivery and come to function
39 as conditioned reinforcers (CR), by acquiring motivational salience through repeated pairing with the
40 drug [72,84]. Although numerous studies demonstrate the fundamental role of CR in producing and
41 maintaining drug-seeking behaviors [72,75,85], how resistant habitual behaviors are to changes in CR
42 remains relatively unexplored. More generally, the fundamental role of Pavlovian cues in the control

1 of reward-seeking behaviors remains largely overlooked in tasks employing self-paced free-operant
2 schedules in absence of conditioned and discriminative stimuli.

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

8 **b. facing the complexity of interactions between decision-making processes.**

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

26
27 Evidence of a shift from ventro-medial to dorso-lateral striatum in striato-nigro-striatal dopaminergic
28 pathways, which is proposed to underlie the transition from goal-directed to habitual control over
29 drug seeking remains limited. Indeed, studies demonstrating this shift during cocaine self-
30 administration under a second-order schedule of reinforcement did not assess whether behavior was
31 habitual [92]. Although a shift from ventromedial to dorsolateral striatal (DLS) dopamine release has
32 been observed during cocaine self-administration, this shift was suggested to promote refinement of
33 instrumental learning rather than escalated and compulsive cocaine seeking [93]. Numerous studies
34 suggest that DMS and DLS are sequentially involved during early and late instrumental training, when
35 behavior is goal-directed or habitual, respectively [94–97]. This dissociation between DMS and DLS
36 has also been reported following ethanol and cocaine self-administration [11,24]. Furthermore,
37 dopamine transmission in the DMS and DLS is required for early and late performance of cue-
38 mediated cocaine seeking, respectively [98]. However, the hypothesis of sequential involvement of
39 DMS and DLS across habitual learning has been recently challenged [56] and whether this serial
40 recruitment in dorso-striatal activity is accelerated by drug exposure remains unknown. Clearly, more
41 research is needed to demonstrate a shift in meso-nigro-striatal dopaminergic signaling and
42 dorso-striatal activity in the context of habitual drug-seeking behavior.

43

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

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

31 One key problem of goal-directed, model-based strategy is the high computational demand for
32 implementation. In theory, to make decisions under model-based control, agents build a decision
33 tree of all possible states and actions and navigate in this “cognitive map” to estimate the long-run
34 worth of each available outcome [48]. In the forest of decision-tree possibilities in real-world
35 settings, considering all the available options is not possible; relevant paths must be somehow
36 preselected [116]. For instance, possible outcomes in a choice situation may be irrelevant and not
37 considered in the first place. We have recently shown in rats that options can be available but not
38 considered in the associative structure of the task, despite the engagement of goal-directed control
39 [117]. In this task, we allowed rats to exert goal-directed control over the occurrence of choice trials
40 by requiring them to nosepoke in a hole for the presentation of cocaine and saccharin levers (Figure
41 3A). As expected, we found that rats preferred saccharin over cocaine but intriguingly, this
42 preference was exclusive in the majority of rats (Figure 3B). When the interest for saccharin was
43 temporarily lost due to repeated choice (i.e. specific satiety), rats preferred to pause for long periods
44 before reinitiating a choice trial for saccharin, instead of switching to cocaine (Figure 3C). To explain

1 this suboptimal behavior, we suggested that rats are preferentially associating the initiation of
2 behavioral sequences with saccharin, thereby ignoring the drug reward. These results show that in
3 some situations, choice outcomes can be available but ignored, even when responding is under goal-
4 directed control [117].

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

25 **Conclusion**

26 We hope it is clear from this review that habits alone cannot account for the development of
27 compulsive drug use and that drug habits are not necessary [125], nor sufficient [89] to explain the
28 transition to addiction. However, this does not preclude a role for habits in addiction. Then, to what
29 extent are drug habits actually involved? To answer this question, we suffer from several limitations.
30 The structure of our procedures generally favors reengagement of goal-directed control precluding
31 correct assessment of habit. Experiments in animals suffer from a paucity of reward-predictive cues,
32 which does not reflect the sensorial and associative richness of drug addicts' environment and does
33 not facilitate the development of habit by reducing reinforcement uncertainty. Finally, investigations
34 are limited by too narrow views that drug-seeking behavior should be either habitual or goal-
35 directed. Moving forward, we propose to better design instrumental tasks, in the presence of choice
36 and reward-predictive cues, and under conditions of high reinforcement predictability to favor
37 implementation of simple stimulus-response model-free policies. Alternative task structures
38 involving sequential rather than simultaneous choice should also be considered. On a theoretical
39 level, we may need to consider a more complex framework taking into account (1) the continuous
40 arbitration between goal-directed and habitual systems, (2) the hierarchical decision-making
41 architectures combining these two systems and (3) alternative sequential decision-making models
42 suggesting that individuals may consider one option at a time when making decisions. Although
43 much remains to be done, our hope is that this review opens up new perspectives to determine the
44 role of habit and choice in addiction.

1 **Acknowledgement**

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

5 **Funding and Disclosure**

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

10 **Authors' contribution**

11 YV drafted the first version of the manuscript; SHA and YV revised and edited the manuscript; YV and
12 SHA approved the final version of the manuscript.

13 **Reference**

- 14 1. Redish AD, Jensen S, Johnson A. A unified framework for addiction: vulnerabilities in the
15 decision process. *Behav Brain Sci.* 2008;31:415-487.
- 16 2. Everitt BJ, Robbins TW. Drug Addiction: Updating Actions to Habits to Compulsions Ten Years
17 On. *Annu Rev Psychol.* 2016;67:23-50.
- 18 3. Everitt BJ, Robbins TW. Neural systems of reinforcement for drug addiction: from actions to
19 habits to compulsion. *Nat Neurosci.* 2005;8:1481-1489.
- 20 4. Tiffany ST. A cognitive model of drug urges and drug-use behavior: role of automatic and
21 nonautomatic processes. *Psychol Rev.* 1990;97:147-168.
- 22 5. Heather N. Is the concept of compulsion useful in the explanation or description of addictive
23 behaviour and experience? *Addict Behav Reports.* 2017;6:15-38.
- 24 6. Ostlund SB, Balleine BW. On habits and addiction: An associative analysis of compulsive drug
25 seeking. *Drug Discov Today Dis Model.* 2008;5:235-245.
- 26 7. Dickinson A, Balleine B. Motivational Control of Instrumental Action. *Anim Learn Behav.*
27 1994;22:1-18.
- 28 8. Balleine BW, Dickinson A. Goal-directed instrumental action: Contingency and incentive
29 learning and their cortical substrates. *Neuropharmacology.* 1998;37:407-419.
- 30 9. Hogarth L, Lam-Cassettari C, Pacitti H, Currah T, Mahlberg J, Hartley L, et al. Intact goal-
31 directed control in treatment-seeking drug users indexed by outcome-devaluation and
32 Pavlovian to instrumental transfer: Critique of habit theory. *Eur J Neurosci.* 2018:1-13.
- 33 10. Hogarth L. Addiction is driven by excessive goal-directed drug choice under negative affect:
34 translational critique of habit and compulsion theory. *Neuropsychopharmacology.* 2020.
35 2020. <https://doi.org/10.1038/s41386-020-0600-8>.
- 36 11. Corbit L, Nie H, Janak P. Habitual alcohol seeking: time course and the contribution of
37 subregions of the dorsal striatum. *Biol Psychiatry.* 2012;72:389-395.

- 1 12. Barker JM, Zhang H, Villafane JJ, Wang TL, Torregrossa MM, Taylor JR. Epigenetic and
2 pharmacological regulation of 5HT3 receptors controls compulsive ethanol seeking in mice.
3 Eur J Neurosci. 2014;39:999-1008.
- 4 13. Dickinson A, Wood N, Smith JW. Alcohol seeking by rats: action or habit? Q J Exp Psychol B.
5 2002;55:331-348.
- 6 14. Lopez MF, Becker HC, Chandler LJ. Repeated episodes of chronic intermittent ethanol
7 promote insensitivity to devaluation of the reinforcing effect of ethanol. Alcohol. 2014;48:639
8 -645.
- 9 15. Mangieri RA, Cofresí RU, Gonzales RA. Ethanol seeking by long evans rats is not always a goal-
10 directed behavior. PLoS One. 2012;7:1-13.
- 11 16. Mangieri R a, Cofresí RU, Gonzales R a. Ethanol exposure interacts with training conditions to
12 influence behavioral adaptation to a negative instrumental contingency. Front Behav
13 Neurosci. 2014;8:220.
- 14 17. Corbit LH, Nie H, Janak PH. Habitual responding for alcohol depends upon both AMPA and D2
15 receptor signaling in the dorsolateral striatum. Front Behav Neurosci. 2014;8.
- 16 18. Miles FJ, Everitt BJ, Dickinson A. Oral cocaine seeking by rats: action or habit? Behav Neurosci.
17 2003;117:927-938.
- 18 19. Leong KC, Berini CR, Ghee SM, Reichel CM. Extended cocaine-seeking produces a shift from
19 goal-directed to habitual responding in rats. Physiol Behav. 2016;164:330-335.
- 20 20. Clemens KJ, Lay BP, Holmes NM. Extended nicotine self-administration increases sensitivity to
21 nicotine, motivation to seek nicotine and the reinforcing properties of nicotine-paired cues.
22 Addict Biol. 2015. 2015.
- 23 21. Loughlin A, Funk D, Coen K, Lê AD. Habitual nicotine-seeking in rats following limited training.
24 Psychopharmacology (Berl). 2017:1-11.
- 25 22. Olmstead MC, Lafonda M V., Everittb BJ, Dickinsonb A. Cocaine seeking by rats is a goal-
26 directed action. Behav Neurosci. 2001;115:394-402.
- 27 23. Hutcheson DM, Everitt BJ, Robbins TW, Dickinson A. The role of withdrawal in heroin
28 addiction: Enhances reward or promotes avoidance? Nat Neurosci. 2001;4:943-947.
- 29 24. Zapata A, Minney VL, Shippenberg TS. Shift from goal-directed to habitual cocaine seeking
30 after prolonged experience in rats. J Neurosci. 2010;30:15457-15463.
- 31 25. Renteria R, Baltz ET, Gremel CM. Chronic alcohol exposure disrupts top-down control over
32 basal ganglia action selection to produce habits. Nat Commun. 2018;9:1-11.
- 33 26. Corbit LH, Chieng BC, Balleine BW. Effects of repeated cocaine exposure on habit learning and
34 reversal by N-acetylcysteine. Neuropsychopharmacology. 2014;39:1893-1901.
- 35 27. LeBlanc KH, Maidment NT, Ostlund SB. Repeated Cocaine Exposure Facilitates the Expression
36 of Incentive Motivation and Induces Habitual Control in Rats. PLoS One. 2013;8.
- 37 28. Nelson A, Killcross S. Amphetamine Exposure Enhances Habit Formation. J Neurosci.
38 2006;26:3805-3812.

- 1 29. Nordquist RE, Voorn P, de Mooij-van Malsen JG, Joosten RNJMA, Pennartz CMA,
2 Vanderschuren LJMJ. Augmented reinforcer value and accelerated habit formation after
3 repeated amphetamine treatment. *Eur Neuropsychopharmacol.* 2007;17:532-540.
- 4 30. Nelson AJD, Killcross S, Leblanc KH. Accelerated habit formation following amphetamine
5 exposure is reversed by D 1 , but enhanced by D 2 , receptor antagonists. *Front Neurosci.*
6 2013;7:1-13.
- 7 31. Schmitzer-Torbert N, Apostolidis S, Amoa R, O'Rear C, Kaster M, Stowers J, et al. Post-training
8 cocaine administration facilitates habit learning and requires the infralimbic cortex and
9 dorsolateral striatum. *Neurobiol Learn Mem.* 2015;118:105-112.
- 10 32. Shiflett MW. The effects of amphetamine exposure on outcome-selective Pavlovian-
11 instrumental transfer in rats. *Psychopharmacology (Berl).* 2012;223:361-370.
- 12 33. Gourley SL, Olevska A, Gordon J, Taylor JR. Cytoskeletal Determinants of Stimulus-Response
13 Habits. *J Neurosci.* 2013;33:11811-11816.
- 14 34. Halbout B, Liu AT, Ostlund SB. A closer look at the effects of repeated cocaine exposure on
15 adaptive decision-making under conditions that promote goal-directed control. *Front*
16 *Psychiatry.* 2016;7:1-12.
- 17 35. Phillips GD, Vugler A. Effects of sensitization on the detection of an instrumental contingency.
18 *Pharmacol Biochem Behav.* 2011;100:48-58.
- 19 36. Son JH, Latimer C, Keefe KA. Impaired formation of stimulus-response, but not action-
20 outcome, associations in rats with methamphetamine-induced neurotoxicity.
21 *Neuropsychopharmacology.* 2011;36:2441-2451.
- 22 37. Dickinson A, Nicholas DJ, Adams CD. The effect of the instrumental training contingency on
23 susceptibility to reinforcer devaluation. *Q J Exp Psychol Sect B.* 1983;35:35-51.
- 24 38. Derusso AL, Fan D, Gupta J, Shelest O, Costa RM, Yin HH. Instrumental uncertainty as a
25 determinant of behavior under interval schedules of reinforcement. *Front Integr Neurosci.*
26 2010;4:1-8.
- 27 39. Urcelay GP, Jonkman S. Delayed Rewards Facilitate Habit Formation Delayed Rewards
28 Facilitate Habit Formation. *J Exp Psychol Anim Learn Cogn.* 2019;August 1.
- 29 40. Adams CD. Variations in the sensitivity of instrumental responding to reinforcer devaluation.
30 *Q J Exp Psychol Sect B.* 1982;34:77-98.
- 31 41. Dickinson A. Actions and Habits: The Development of Behavioural Autonomy. *Philos Trans R*
32 *Soc B Biol Sci.* 1985;308:67-78.
- 33 42. Holland PC. Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *J*
34 *Exp Psychol Anim Behav Process.* 2004;30:104-117.
- 35 43. Colwill RM, Triola SM. Instrumental responding remains under the control of the consequent
36 outcome after extended training. *Behav Processes.* 2002;57:51-64.
- 37 44. Kosaki Y, Dickinson A. Choice and contingency in the development of behavioral autonomy
38 during instrumental conditioning. *J Exp Psychol Anim Behav Process.* 2010;36:334-342.
- 39 45. Trask S, Shipman ML, Green JT, Bouton ME. Some factors that restore goal-direction to a

- 1 habitual behavior. *Neurobiol Learn Mem.* 2020;169.
- 2 46. Bouton ME, Broomer MC, Rey CN, Thrailkill EA. Unexpected food outcomes can return a habit
3 to goal-directed action. *Neurobiol Learn Mem.* 2020;169:1-9.
- 4 47. Thrailkill EA, Trask S, Vidal P, Alcalá JA, Bouton ME. Stimulus Control of Actions and Habits : A
5 Role for Reinforcer Predictability and Attention in the Development of Habitual Behavior. *J*
6 *Exp Psychol Anim Learn Cogn.* 2018;44:370-384.
- 7 48. Dolan RJ, Dayan P. Goals and habits in the brain. *Neuron.* 2013;80:312-325.
- 8 49. Keramati M, Smittenaar P, Dolan RJ, Dayan P. Adaptive integration of habits into depth-
9 limited planning defines a habitual-goal – directed spectrum. *Proc Natl Acad Sci U S A.*
10 2016;113:12868-12873.
- 11 50. Vandaele Y, Guillem K, Ahmed SH, Ahmed SH. Habitual Preference for the Nondrug Reward in
12 a Drug Choice Setting. *Front Behav Neurosci.* 2020;14:1-9.
- 13 51. Vandaele Y, Vouillac-Mendoza C, Ahmed SH. Inflexible habitual decision-making during choice
14 between cocaine and a nondrug alternative. *Transl Psychiatry.* 2019;9.
- 15 52. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral
16 striatal systems for behavioral control. *Nat Neurosci.* 2005;8:1704-1711.
- 17 53. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans'
18 choices and striatal prediction errors. *Neuron.* 2011;69:1204-1215.
- 19 54. Doya K, Samejima K, Katagiri K, Kawato M. Multiple Model-Based Reinforcement Learning.
20 *Neural Comput.* 2002;14:1347-1369.
- 21 55. Vandaele Y, Pribut HJ, Janak PH. Lever Insertion as a Salient Stimulus Promoting Insensitivity
22 to Outcome Devaluation. *Front Integr Neurosci.* 2017;11.
- 23 56. Vandaele Y, Mahajan NR, Ottenheimer DJ, Richard JM, Mysore SP, Janak PH. Distinct
24 recruitment of dorsomedial and dorsolateral striatum erodes with extended training. *Elife.*
25 2019;8:1-29.
- 26 57. Lee SW, Shimojo S, O'Doherty JP. Neural Computations Underlying Arbitration between
27 Model-Based and Model-free Learning. *Neuron.* 2014;81:687-699.
- 28 58. Colwill RM, Rescorla RA. Instrumental responding remains sensitive to reinforcer devaluation
29 after extensive training. *J Exp Psychol Anim Behav Process.* 1985;11:520-536.
- 30 59. Parkes SL, Balleine BW. Incentive Memory : Evidence the Basolateral Amygdala Encodes and
31 the Insular Cortex Retrieves Outcome Values to Guide Choice between Goal-Directed Actions.
32 *J Neurosci.* 2013;33:8753-8763.
- 33 60. Parkes SL, Bradfield LA, Balleine BW. Interaction of Insular Cortex and Ventral Striatum
34 Mediates the Effect of Incentive Memory on Choice Between Goal-Directed Actions. *J*
35 *Neurosci.* 2015;35:6464-6471.
- 36 61. Balleine BW, Killcross AS, Dickinson A. The Effect of Lesions of the Basolateral Amygdala on
37 Instrumental Conditioning. *J Neurosci.* 2003;23:666-675.
- 38 62. Corbit LH, Balleine BW. The role of prelimbic cortex in instrumental conditioning. *Behav Brain*

- 1 Res. 2003;146:145-157.
- 2 63. Glimcher PW, Rustichini A. Neuroeconomics: The consilience of brain and decision. *Science*
3 (80-). 2004;306:447-452.
- 4 64. Rangel A, Camerer C, Montague PR. A framework for studying the neurobiology of value-
5 based decision making. *Nat Rev Neurosci*. 2008;9:545-556.
- 6 65. Rushworth MF, Mars RB, Summerfield C. General mechanisms for making decisions? *Curr*
7 *Opin Neurobiol*. 2009;19:75-83.
- 8 66. Rangel A, Hare T. Neural computations associated with goal-directed choice. *Curr Opin*
9 *Neurobiol*. 2010;20:262-270.
- 10 67. Keramati M, Dezfouli A, Piray P. Speed/Accuracy Trade-Off between the Habitual and the
11 Goal-Directed Processes. *PLoS Comput Biol*. 2011;7:e1002055.
- 12 68. Lenoir M, Augier E, Vouillac C, Ahmed SH. A Choice-based screening method for compulsive
13 drug users in rats. *Curr Protoc Neurosci*. 2013;1:1-17.
- 14 69. Cantin L, Lenoir M, Augier E, Vanhille N, Dubreucq S, Serre F, et al. Cocaine is low on the value
15 ladder of rats: Possible evidence for resilience to addiction. *PLoS One*. 2010;5.
- 16 70. Lenoir M, Serre F, Cantin L, Ahmed SH. Intense Sweetness Surpasses Cocaine Reward. *PLoS*
17 *One*. 2007;2.
- 18 71. Ahmed SH. The science of making drug-addicted animals. *Neuroscience*. 2012;211:107-125.
- 19 72. Namba MD, Tomek SE, Olive MF, Beckmann JS, Gipson CD. The winding road to relapse:
20 Forging a new understanding of cue-induced reinstatement models and their associated
21 neural mechanisms. *Front Behav Neurosci*. 2018;12:1-22.
- 22 73. Weiss F, Maldonado-Vlaar CS, Parsons LH, Kerr TM, Smith DL, Ben-Shahar O. Control of
23 cocaine-seeking behavior by drug-associated stimuli in rats: Effects on recovery of
24 extinguished operant-responding and extracellular dopamine levels in amygdala and nucleus
25 accumbens. *Proc Natl Acad Sci U S A*. 2000;97:4321-4326.
- 26 74. Weiss F, Martin-Fardon R, Ciccocioppo R, Kerr TM, Smith DL, Ben-Shahar O. Enduring
27 resistance to extinction of cocaine-seeking behavior induced by drug-related cues.
28 *Neuropsychopharmacology*. 2001;25:361-372.
- 29 75. Shaham Y, Shalev U, Lu L, de Wit H, Stewart J. The reinstatement model of drug relapse:
30 history, methodology and major findings. *Psychopharmacology (Berl)*. 2003;168:3-20.
- 31 76. Crombag HS, Bossert JM, Koya E, Shaham Y. Context-induced relapse to drug seeking: a
32 review. *Philos Trans R Soc B Biol Sci*. 2008;363:3233-3243.
- 33 77. Hogarth L, Chase HW. Parallel goal-directed and habitual control of human drug-seeking:
34 implications for dependence vulnerability. *J Exp Psychol Anim Behav Process*. 2011;37:261-
35 276.
- 36 78. Hogarth L. Goal-directed and transfer-cue-elicited drug-seeking are dissociated by
37 pharmacotherapy: evidence for independent additive controllers. *J Exp Psychol Anim Behav*
38 *Process*. 2012;38:266-278.

- 1 79. Corbit LH, Janak PH, Balleine BW. General and outcome-specific forms of Pavlovian-
2 instrumental transfer: the effect of shifts in motivational state and inactivation of the ventral
3 tegmental area. *Eur J Neurosci.* 2007;26:3141-3149.
- 4 80. Rescorla R a. Transfer of instrumental control mediated by a devalued outcome. *Anim Learn
5 Behav.* 1994;22:27-33.
- 6 81. Watson P, Wiers RW, Hommel B, de Wit S. Working for food you don't desire. Cues interfere
7 with goal-directed food-seeking. *Appetite.* 2014;79:139-148.
- 8 82. Steenbergen H Van, Watson P, Wiers RW, Hommel B, de Wit S. Dissociable corticostriatal
9 circuits underlie goal-directed vs . cue-elicited habitual food seeking after satiation : evidence
10 from a multimodal MRI study. *Eur J Neurosci.* 2017:1-13.
- 11 83. Lamb RJ, Schindler W, Pinkston JW. Conditioned Stimuli's Role in Relapse: Pre-Clinical
12 Research on Pavlovian Instrumental Transfer. *Psychopharmacology (Berl).* 2016;233:1933-
13 1944.
- 14 84. KELLEHER RT, GOLLUB LR. A review of positive conditioned reinforcement. *J Exp Anal Behav.*
15 1962;5:543-597.
- 16 85. Bossert JM, Marchant NJ, Calu DJ, Shaham Y. The reinstatement model of drug relapse: recent
17 neurobiological findings, emerging research topics, and translational research.
18 *Psychopharmacology (Berl).* 2013;229:453-476.
- 19 86. Hu Y, Salmeron BJ, Krasnova IN, Gu H, Lu H, Bonci A, et al. Compulsive drug use is associated
20 with imbalance of orbitofrontal- And prelimbic-striatal circuits in punishment-resistant
21 individuals. *Proc Natl Acad Sci U S A.* 2019;116:9066-9071.
- 22 87. Pascoli V, Terrier J, Hiver A, Lüscher C. Sufficiency of Mesolimbic Dopamine Neuron
23 Stimulation for the Progression to Addiction. *Neuron.* 2015;88:1054-1066.
- 24 88. Pascoli V, Hiver A, Van Zessen R, Loureiro M, Achargui R, Harada M, et al. Stochastic synaptic
25 plasticity underlying compulsion in a model of addiction. *Nature.* 2018;564:366-371.
- 26 89. Lüscher C, Robbins TW, Everitt BJ. The transition to compulsion in addiction. *Nat Rev
27 Neurosci.* 2020;21:247-263.
- 28 90. Goldstein RZ, Volkow ND. Dysfunction of the prefrontal cortex in addiction: neuroimaging
29 findings and clinical implications. *Nat Rev Neurosci.* 2012;12:652-669.
- 30 91. Baler RD, Volkow ND. Drug addiction: the neurobiology of disrupted self-control. *Trends Mol
31 Med.* 2006;12:559-566.
- 32 92. Belin D, Everitt BJ. Cocaine Seeking Habits Depend upon Dopamine-Dependent Serial
33 Connectivity Linking the Ventral with the Dorsal Striatum. *Neuron.* 2008;57:432-441.
- 34 93. Willuhn I, Burgeno LM, Everitt BJ, Phillips PEM. Hierarchical recruitment of phasic dopamine
35 signaling in the striatum during the progression of cocaine use. *Proc Natl Acad Sci U S A.*
36 2012;109:20703-20708.
- 37 94. Yin HH, Ostlund SB, Knowlton BJ, Balleine BW. The role of the dorsomedial striatum in
38 instrumental conditioning. *Eur J Neurosci.* 2005;22:513-523.
- 39 95. Yin HH, Knowlton BJ, Balleine BW. Lesions of dorsolateral striatum preserve outcome

- 1 expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci*. 2004;19:181-
2 189.
- 3 96. Yin HH, Knowlton BJ. The role of the basal ganglia in habit formation. *Nat Rev Neurosci*.
4 2006;7:464-476.
- 5 97. Balleine BW, Liljeholm M, Ostlund SB. The integrative function of the basal ganglia in
6 instrumental conditioning. *Behav Brain Res*. 2009;199:43-52.
- 7 98. Murray JE, Belin D, Everitt BJ. Double dissociation of the dorsomedial and dorsolateral striatal
8 control over the acquisition and performance of cocaine seeking. *Neuropsychopharmacology*.
9 2012;37:2456-2466.
- 10 99. Preble E. Taking Care of Business- The Heroin User ' s Life on the Street *. *Int J Addict*.
11 1969;4:1-24.
- 12 100. Schreiner DC, Renteria R, Gremel CM. Fractionating the all-or-nothing definition of goal-
13 directed and habitual decision-making. *J Neurosci Res*. 2019:1-9.
- 14 101. Feher da Silva C, Hare TA. Humans primarily use model-based inference in the two-stage task.
15 *Nat Hum Behav*. 2020. juillet 2020. <https://doi.org/10.1038/s41562-020-0905-y>.
- 16 102. Byrne KA, Otto AR, Pang B, Patrick CJ, Worthy DA. Substance use is associated with reduced
17 devaluation sensitivity. *Cogn Affect Behav Neurosci*. 2019;19:40-55.
- 18 103. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND. Characterizing a psychiatric symptom
19 dimension related to deficits in goal-directed control. *Elife*. 2016;5.
- 20 104. Miller KJ, Botvinick MM, Brody CD. Dorsal hippocampus contributes to model-based planning.
21 *Nat Neurosci*. 2017;20:1269-1276.
- 22 105. Groman SM, Massi B, Mathias SR, Lee D, Taylor JR. Model-Free and Model-Based Influences in
23 Addiction-Related Behaviors. *Biol Psychiatry*. 2019;85:936-945.
- 24 106. Groman SM, Massi B, Mathias SR, Curry DW, Lee D, Taylor JR. Neurochemical and behavioral
25 dissections of decision-making in a rodent multistage task. *J Neurosci*. 2019;39:295-306.
- 26 107. Fraser KM, Janak PH. How Does Drug Use Shift the Balance Between Model-Based and Model-
27 Free Control of Decision Making? *Biol Psychiatry*. 2019;85:886-888.
- 28 108. Otto AR, Gershman SJ, Markman AB, Daw ND. The Curse of Planning. *Psychol Sci*. 2013;24:751
29 -761.
- 30 109. Dezfouli A, Balleine BW. Actions, action sequences and habits: evidence that goal-directed
31 and habitual action control are hierarchically organized. *PLoS Comput Biol*. 2013;9:e1003364.
- 32 110. Dezfouli A, Balleine BW. Habits, action sequences and reinforcement learning. *Eur J Neurosci*
33 2012 Apr;35(7)1036-51 doi 10.1111/j.1460-9568.2012.08050.x. 2012. 2012.
- 34 111. Dezfouli A, Lingawi NW, Balleine BW. Habits as action sequences: hierarchical action control
35 and changes in outcome value. *Philos Trans R Soc L B Biol Sci*. 2014;369.
- 36 112. Balleine BW, Dezfouli A. Hierarchical Action Control: Adaptive Collaboration Between Actions
37 and Habits. *Front Psychol*. 2019;10.

- 1 113. Cushman F, Morris A. Habitual control of goal selection in humans. *Proc Natl Acad Sci*.
2 2015;112:13817-13822.
- 3 114. Wood W, Neal DT. A New Look at Habits and the Habit – Goal Interface. *Psychol Rev*.
4 2007;114:843-863.
- 5 115. Kruglanski AW, Szumowska E. Habitual Behavior Is Goal-Driven. *Perspect Psychol Sci*.
6 2020:174569162091767.
- 7 116. Huys QJM, Eshel N, O’Nions E, Sheridan L, Dayan P, Roiser JP. Bonsai trees in your head: How
8 the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput*
9 *Biol*. 2012;8.
- 10 117. Vandaele Y, Vouillac-Mendoza C, Ahmed SH. Cocaine falls into oblivion during volitional
11 initiation of choice trials. *Addict Biol*. 2019:in press.
- 12 118. Shapiro MS, Siller S, Kacelnik A. Simultaneous and Sequential Choice as a Function of Reward
13 Delay and Magnitude: Normative, Descriptive and Process-Based Models Tested in the
14 European Starling (*Sturnus vulgaris*). *J Exp Psychol Anim Behav Process*. 2008;34:75-93.
- 15 119. Freidin E, Aw J, Kacelnik A. Sequential and simultaneous choices: Testing the diet selection
16 and sequential choice models. *Behav Processes*. 2009;80:218-223.
- 17 120. Freidin E, Kacelnik A. Rational choice, context dependence, and the value of information in
18 European starlings (*Sturnus vulgaris*). *Science* (80-). 2011;334:1000-1002.
- 19 121. Vasconcelos M, Monteiro T, Aw J, Kacelnik A. Choice in multi-alternative environments: A
20 trial-by-trial implementation of the Sequential Choice Model. *Behav Processes*. 2010;84:435-
21 439.
- 22 122. Vasconcelos M, Monteiro T, Kacelnik A. Context-Dependent Preferences in Starlings: Linking
23 Ecology, Foraging and Choice. *PLoS One*. 2013;8.
- 24 123. Mobbs D, Trimmer PC, Blumstein DT, Dayan P. Foraging for foundations in decision
25 neuroscience: insights from ethology. *Nat Rev Neurosci*. 2018;19:419-427.
- 26 124. Grace RC. Acquisition of choice in concurrent chains: Assessing the cumulative decision
27 model. *Behav Processes*. 2016;126:82-93.
- 28 125. Singer BF, Fadanelli M, Kawa AB, Robinson TE. Are cocaine-seeking “ habits ” necessary for the
29 development of addiction-like behavior in rats ? *J Neurosci*. 2017;38:60-73.

30

31

1 **Figure legends**

2 **Figure 1: habitual preference for saccharin in a drug choice setting.** A-C. Responding for saccharin is not
3 reduced following saccharin devaluation by specific satiety. A. Rats' performance on the cocaine and saccharin
4 levers did not differ between the Devalued group (D; white) and the Non-Devalued group (ND; black) across 1-
5 min time bins in the extinction test. * $p < 0.05$ Coc vs Sacch. B. The total number of lever presses was higher on
6 the saccharin lever compared to the cocaine lever but was not affected by devaluation. * $p < 0.05$ Coc vs Sacch.
7 C. Saccharin was correctly devalued as measured by a reduction in post-test consumption of saccharin in the D
8 group compared to the ND group. D-F. Preference for saccharin is also insensitive to saccharin devaluation by
9 CTA. D-E. Rats responded more on the saccharin lever compared to the cocaine lever but did not differ as a
10 function of devaluation. * $p < 0.05$ Coc vs Sacch. F. Devaluation of saccharin was confirmed during the test of
11 consumption immediately after the extinction session. Adapted from [50]

12 **Figure 2: Inflexible preference for the alternative non-drug reward in a drug choice setting is under habitual,**
13 **model-free control.** Water-restricted rats offered a choice between water and cocaine expressed a robust
14 preference for water (black; baseline preference under water deprivation). Water was then partially devalued
15 with 1h (1h- \emptyset , pink) and 2h free water access (2h- \emptyset , purple) before the choice session. Water preference was
16 not affected (A) but there was moderate suppression of water consumption. ~~n was moderate~~ (B). Thus, free-
17 water access was also introduced during each inter-trial interval (ITI) of choice sessions in addition to the hour
18 of water pre-session access (white; 1h+ITI, Free Water FW). Although this condition drastically suppressed
19 water consumption from the first FW session (B), 9 sessions were needed to observe a complete reversal of
20 preference (A). Following this devaluation training, 1h water access was sufficient to raise cocaine preference
21 to 50% in a second 1h- \emptyset choice session (pink). Finally, devaluation of water by taste adulteration with quinine
22 only moderately affected preference (A) despite a strong suppression of water consumption (B). Adapted from
23 [51]

24 **Figure 3: Rats are oblivious to the cocaine option during self-initiated choice.** (A) Rats are required to
25 nosepoke in a hole under a fixed ratio 10 to trigger the presentation of two levers. Two consecutive presses on
26 the left or right lever result in the delivery of saccharin or an intravenous infusion of cocaine, respectively. (B)
27 In this procedure, rats expressed a strong preference for saccharin. Interestingly, this preference was exclusive
28 for a majority of rats (right panel). (C) Analysis of choice patterns reveals that rats choosing saccharin
29 exclusively did so in bouts of varying lengths separated by pauses, during which they did not self-initiated any
30 trial for cocaine, despite transient saccharin devaluation by sensory-specific satiety. This behavior represents
31 an opportunity cost because the duration of pauses is sufficient to earn several cocaine injections (right panel).
32 Adapted from (Vandaele et al., 2019c).

33

1 Boxes

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, 1998; Dickinson, 1985, 1994).

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 a conditioned taste aversion (Adams and Dickinson, 1981; Balleine and Dickinson, 1998; Colwill, 1993; Dickinson and Balleine, 1994; Rescorla, 1987). 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 non-contingent 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 response should be reduced compared to the non-degraded alternative (Colwill, 1993; Dickinson and Mulatero, 1989). Conversely, insensitivity to this procedure indicates that performance is under habitual control.*

- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16

Box 2 –MODEL-BASED AND MODEL-FREE CONTROL

Algorithms in reinforcement learning, namely model-based (MB) and model-free (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 model-based and model-free control by assessing subjects’ trial-by-trial sensitivity to immediate reward history and task structure (Daw et al., 2011; Gläscher et al., 2010). 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 model-based 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 model-based agent will switch to the alternative 1st-step option (more likely to result in the previously rewarded 2nd-step state) whereas the model-free agent will repeat the same 1st-stage choice with no adjustment based on the type of transition.

