Review

Biological Psychiatry

Model-Based Control in Dimensional Psychiatry

Valerie Voon, Andrea Reiter, Miriam Sebold, and Stephanie Groman

ABSTRACT

We use parallel interacting goal-directed and habitual strategies to make our daily decisions. The arbitration between these strategies is relevant to inflexible repetitive behaviors in psychiatric disorders. Goal-directed control, also known as model-based control, is based on an affective outcome relying on a learned internal model to prospectively make decisions. In contrast, habit control, also known as model-free control, is based on an integration of previous reinforced learning autonomous of the current outcome value and is implicit and more efficient but at the cost of greater inflexibility. The concept of model-based control can be further extended into pavlovian processes. Here we describe and compare tasks that tap into these constructs and emphasize the clinical relevance and translation of these tasks in psychiatric disorders. Together, these findings highlight a role for model-based control as a transdiagnostic impairment underlying compulsive behaviors and representing a promising therapeutic target.

Keywords: Addictions, Binge eating, Compulsivity, Computational psychiatry, Goal-directed control, Habit, Model-based control, Obsessive-compulsive disorder

http://dx.doi.org/10.1016/j.biopsych.2017.04.006

We use parallel interacting goal-directed and habitual strategies to make our daily decisions, both mundane and complex. These decisions include simple ones, such as which road to drive to work, to more complex ones, such as which of multiple options to select as an investment strategy. The capacity to arbitrate between these strategies is relevant to inflexible repetitive behaviors observed in psychiatric disorders and represents a critical construct in dimensional psychiatry.

Model-based control describes a process that relies on a learned internal model of the environment to prospectively evaluate actions based on their potential outcomes (1,2). The associative structure of the model is stored and includes predictions about the consequences of each state and can be used to mentally simulate and infer values and outcomes that go beyond our previous experience. This strategy is effective and flexible, particularly with changing and novel environments, but can be computationally expensive. Goal-directed behaviors are a form of model-based control describing instrumental responding that is sensitive to the contingencies between responses and outcomes and the current value of the outcome. In contrast, model-free control describes a process in which prediction errors (what we actually receive vs. what we expected to receive) are used to estimate and store action values based on past experience. This strategy is more implicit, efficient, and rapid, with decisions based on retrospective stored values but at the cost of greater inflexibility. Habit control is a form of instrumental behavior in which responding persists despite changes in the current outcome value and represents a form of model-free control. The conceptual discrimination between flexible model-based and stored value-based behavior is commonly applied to instrumental processes using multistep tasks (1,2) but can also be applied to pavlovian processes (3). The capacity to arbitrate between these strategies is relevant dimensionally across compulsive behaviors in psychiatry.

Here we emphasize the clinical and translational relevance of goal-directed and habit control in patient populations characterized by a behavioral phenotype of compulsivity. Then we focus on the underlying neural correlates and characteristics of the two-step task highlighting commonalities and differences between humans and rodents, between task types and types of associative control.

MEASURES OF GOAL-DIRECTED AND HABIT CONTROL

Measures of goal-directed and habit control in preclinical and human studies can be divided into two basic forms: conventional overtraining and devaluation tasks (Figure 1) (4,5) and sequential decision tasks (also known as multistep tasks, with the most common being the two-step task) (Figure 2). Here we describe these concepts.

Goal-directed and habitual behaviors are common in daily decisions. We might see (stimulus) and take (response) the same turn-off when driving home (the goal). If, however, after many years of repeating this same activity (overtraining), we move to a different home, not uncommonly we will mistakenly take this old turn-off based on a habitual (overlearned stimulus–response) strategy, not taking into account the current value of the outcome (the now devalued wrong home).

Goal-directed control is governed by the knowledge of the association between actions and the value of consequences, also known as stimulus–response–outcome associations (Figure 1). With limited training of these associations, rodents

ISSN: 0006-3223

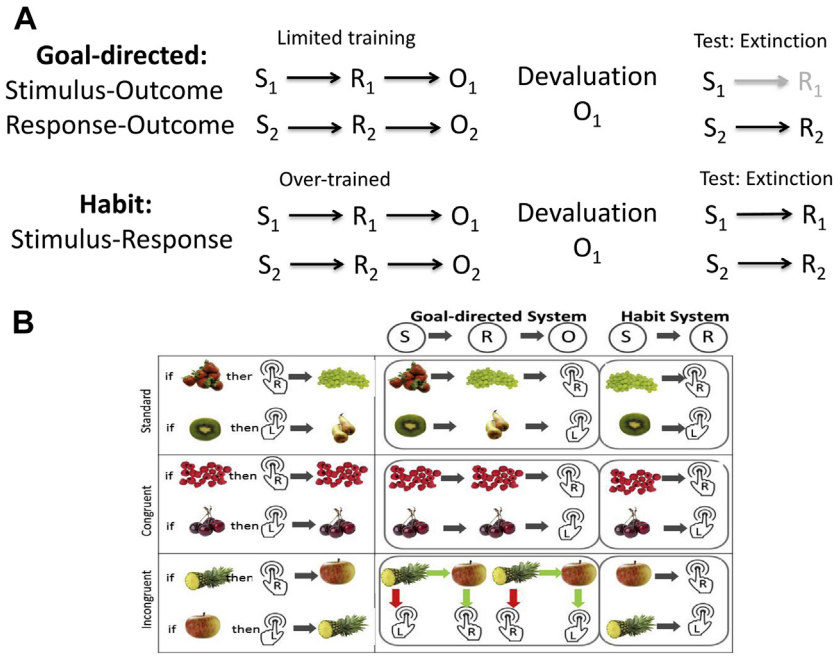

Figure 1. Conventional overtraining and devaluation tasks. **(A)** Rodents and humans undergo training to learn associations between stimulus (S), response (R), and outcome (O) contingencies. Following training, one of the outcomes is devalued (e.g., O_1). Because the subsequent test occurs under extinction (meaning that the devalued outcome is not experienced), the behavior requires access to an internal model of previous learned associations and the current value of the outcome. With moderate training, behavior is guided by stimulus–outcome or response–outcome mappings; hence, responding decreases to the devalued outcome. With extensive training, behavior becomes guided by stimulus–response mappings; hence, responding is autonomous of the current value of the outcome. For example, in this procedure, after rodents learn to obtain two different types of food, one type of food (e.g., O_1) is devalued by pairing with lithium chloride or free access to the food to induce satiety. In the probe test, a decrease in responding is normally observed to the food that is no longer valued (i.e., goal directed), but those with extensive training will persist in responding to the devalued food (i.e., habitual). **(B)** Conflict and slips-of-action task. Subjects first learn the contingencies between six cues (fruit) and responses (left [L] or right [R] button) and outcomes (fruit) for points. [Panel **B** adapted from (6). Images are from open source Stimulus Set (78).] The differentiation between goal-directed and habit learning can be assessed in one of two ways. One way is the congruent and incongruent cue–outcomes test. When the fruit cue and fruit outcome were congruent (i.e., the fruit cue leads to the same fruit outcome), both goal-directed and habitual systems were recruited, whereas only the habitual system was predominantly used when the cue and outcome were incongruent (i.e., the fruit cue leads to a different fruit outcome) because using the goal-directed system would be disadvantageous. The other way is the slips-of-action test (not shown). Instructed outcome devaluation was used to devalue two of the six fruit outcomes (i.e., subjects were told that the outcomes were no longer valuable or associated with loss of points). Subjects were then shown fruit cues in which they could earn points by pressing for the valued fruit outcome or avoid losing points by withholding pressing for the devalued fruit outcome. Habitual slips of action were characterized by pressing for the devalued fruit outcome.

remain sensitive to the current outcome value (i.e., remain goal directed). This can be assessed with devaluation of the outcome and subsequent testing of responding to learned stimuli under extinction conditions when no outcome is present. Because the test occurs without experiencing the outcome, the behavior requires access to an internal model based on previous learned associations and the current value of the outcome. With overtraining of these associations, sensitivity to the current value of the outcome is decreased with increased reliance on stimulus–response associations (i.e., shifts toward habit). Thus, responses become persistent and fail to shift flexibly with changes in current outcome value (4,5).

Human studies have similarly translated overtrained and devaluation tasks. One specific design uses overtraining and testing with a conflict procedure and "slips of action" to assess goal-directed and habit control (6) (Figure 1).

Multistep tasks based on reinforcement learning models have been applied to goal-directed and habit control, also known as model-based and model-free control (1,2,7). The two-step task is a sequential two-stage decision task in which subjects choose between one of two choices at each state, leading in the second stage to a rewarded or nonrewarded outcome of varying probability (1) (Figure 2). Choices at the first stage are associated with a likely transition and an unlikely transition of fixed probability to one of two states. Model-free habitual control is based on the repetition of a previously rewarded action irrespective of the likelihood of the transition, whereas model-based goal-directed control takes into account the task model and the likelihood of the transition. The task provides an index of the relative balance between model-based and model-free control (Figure 2).

NEURAL SUBSTRATES OF GOAL-DIRECTED AND HABIT TASKS IN HEALTHY HUMANS

Rodent and human studies implicate similar dissociable frontostriatal regions in the balance between goal-directed and habit learning. Lesions of the rodent dorsomedial striatum (human caudate) and prelimbic cortex block goal-directed behaviors, leaving intact habit learning (8,9). In contrast, lesions of the dorsolateral striatum (human putamen) and infralimbic cortex result in intact goal-directed behaviors despite extended training (9–11).

Human functional magnetic resonance imaging studies using overtrained and devaluation tasks show a clear dissociation: goal-directed behaviors are associated with ventromedial prefrontal cortex (vmPFC) and caudate activity, regions implicated in action–outcome encoding and outcome valuation relevant to tracking immediate outcome values, and habit learning is associated with putaminal regions. Following training on reinforcement learning tasks, greater habitual behaviors over the course of learning were associated with increased posterior putaminal activity (12) and greater

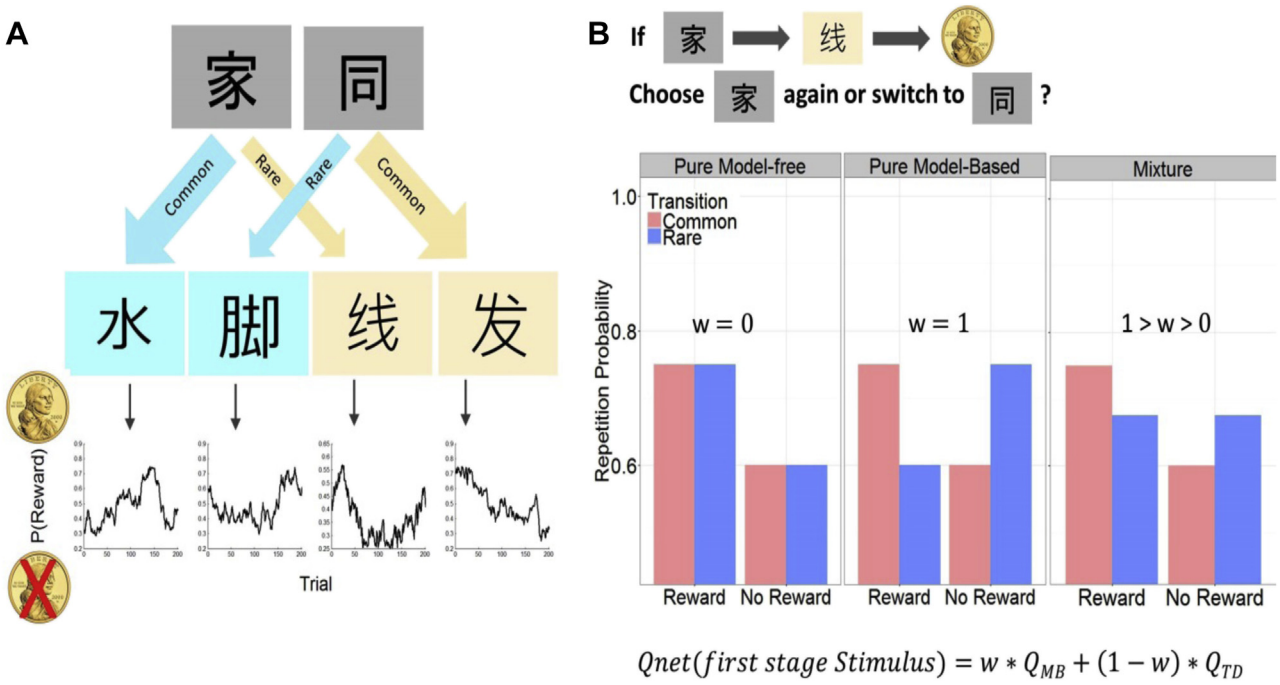

Figure 2. Sequential decision task in humans. **(A)** Example of a two-step trial. Participants first choose between a stimulus pair in the first stage (gray color). This selection then leads to one of two colored second-stage options (blue or yellow color). Again, subjects then choose between a stimulus pair. The transition from first-stage selections to the specific second stage is probabilistic. Each first-stage option leads frequently to one colored second-stage option (Common: 70%) but rarely to the opposing second-stage option (Rare: 30%). After the second-stage selection, participants either are probabilistically rewarded with a monetary reward or do not receive any money. Second-stage reward probabilities (p = .25–.75) change slowly over time according to a random Gaussian walk. **(B)** Expected model-free and model-based response pattern. In the behavioral measure, rare trials allow the dissociation of model-free and model-based control. For instance, in a rare trial, when a first-stage selection unexpectedly leads to a certain second-stage option and this second-stage choice then leads to reward, the model-free agent will stay with the same first-stage stimulus in the subsequent trial and the model-based agent will switch to the opposing first-stage stimulus in the subsequent trial. Thus, model-free first-stage repetitions rely on previously reinforced choices (left panel: main effect of reward), whereas model-based decisions take into account the task structure and transition frequencies from the first stage to the second stage (middle panel: interaction effect of reward and transition). Healthy subjects apply a mixture of both strategies (right panel). In the computational model, parameter values are determined by integrating effects associated with sequences of many choices. The reinforcement learning algorithm shows the hybrid model (mixture: model free and model based) and assumes that first-stage actions are computed according to model-free temporal difference learning (Q_{TD}) and model-based reinforcement learning (Q_{MB}), where the free parameter w weighs between these values.

functional connectivity between the sensorimotor cortex and putamen (13). Similarly, using the conflict task, goal-directed action was associated with greater vmPFC activity (6). Furthermore, habitual slips of action were associated with greater white matter tract strength between the premotor cortex and posterior putamen, whereas goal-directed actions were between the vmPFC and caudate (14).

The neural correlates of outcome prediction error (i.e., the difference between received and expected outcomes) used in model-free updating and of state prediction error (i.e., the discrepancy between the observed and expected state transitions) used in model-based updating can also be dissociated. Model-based control in the two-step task engages higher order associative cortical regions beyond the vmPFC, consistent with rodent prelimbic cortical involvement. State prediction error is represented in the dorsolateral PFC and intraparietal sulcus (2). Furthermore, repetitive transcranial magnetic stimulation to the dorsolateral PFC impairs model-based control, with the left side being dependent on working memory capacity (15). The dorsolateral PFC is implicated in causal learning between cues and outcomes and in sequential action planning, and the intraparietal sulcus is involved in integrating sensory signals to guide and control goal-directed action in space (2). These processes may be particularly relevant to tracking state transition probabilities and the representation of state space (2) and in the interaction among model-based control, working memory, and cognitive control.

Under conditions in which learning of the state transition is emphasized, only reward prediction error was represented in the ventral striatum (2). However, with dynamically changing rewards in the two-step task, which emphasizes a constant trade-off between model-based and model-free control, both forms of prediction error overlap in the ventral striatum and vmPFC, suggesting that the two systems were not necessarily segregated (1). The ventral striatum may provide a substrate for integration of both forms of prediction error. For instance, top-down information from experimenter instructions may influence model-free learning without additional planning (1).

With overtraining, model-based and model-free neural correlates have been further dissociated using a three-step task. Values associated with goal-directed trials were associated with caudate activity, whereas overtrained habit trials were

associated with posterior putamen activity (16). The vmPFC further showed increased connectivity with both caudate and putamen during choice.

The medial orbitofrontal cortex (mOFC), a region involved in outcome representation to guide flexible behaviors, has also been implicated in model-based control. Greater model-based control was associated with greater mOFC gray matter volumes and neurite density (17), extending into vmPFC and the ventral striatum (17), along with greater resting state functional connectivity between mOFC and ventral striatum (18). Greater putaminal neurite complexity was also associated with model-free control (18). Similarly, goal-directed behaviors tested using moderate training and food devaluation have been associated with OFC activity (19). The vmPFC and mOFC in model-based control may reflect different underlying processes. Task-based functional magnetic resonance imaging studies implicating the vmPFC focus on trial-to-trial prediction error and action values consistent with action–outcome encoding in the vmPFC. In contrast, studies implicating the mOFC focus on interindividual differences of the relative balance between strategies and may also capture other cognitive processes relevant to the integration of outcome expectancies with the inferred associative structure of the task. This process might overlap with rodent model-based pavlovian learning, which also implicates the OFC (discussed below).

Taken together, these findings suggest common neural substrates across rodent and human overtrained devaluation tasks. Human sequential learning tasks further extend these substrates (Figure 3). The vmPFC and caudate are implicated in action–outcome encoding relevant to tracking of immediate outcome values. In contrast, the OFC may implicate common mechanisms underlying model-based learning across instrumental and pavlovian tasks involving the integration of inferred values with associative task structure. Model-based control also implicates higher order lateral associative regions relevant to tracking state transitions, state representation, and associative learning. When the trade-off between strategies is enhanced, the two forms of prediction error overlap, emphasizing integration of both forms of control. Critically, with overtraining on the multistep task, model-free control shifts toward putaminal engagement, consistent with rodent studies.

The two-step task developed originally in human subjects has also been back-translated to animals (20–23) with high translational fidelity (Figure 4 and Supplement).

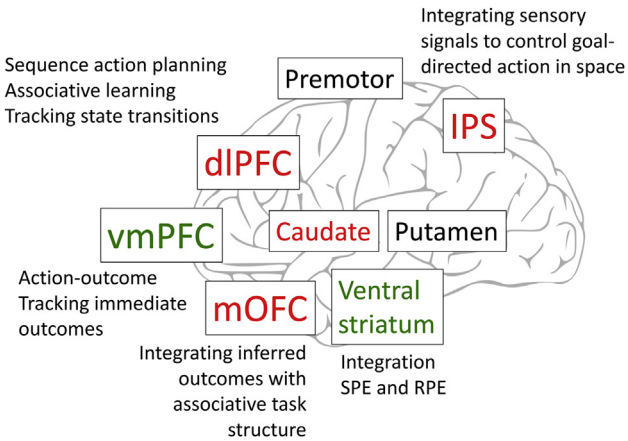

Figure 3. Human neural substrates of model-based and model-free control. Shown is a summary of neural substrates and potential relevant mechanisms implicated in model-based control (in red, dorsolateral prefrontal cortex [dlPFC] and intraparietal sulcus [IPS]), model-free habit control (in black), and overlap of state prediction error (SPE) and reward prediction error (RPE) (in green, ventromedial prefrontal cortex [vmPFC] and ventral striatum). The medial orbitofrontal cortex (mOFC) is implicated in both model-based instrumental and Pavlovian learning.

RELEVANCE TO DIMENSIONAL PSYCHIATRY

The role of model-based control as a dimensional construct underlying compulsive behaviors has been examined across multiple psychiatric disorders. Compulsive behaviors can be defined as a rigid pattern of maladaptive behaviors despite negative consequences (24). We examine similar behavioral phenotypes, namely drug seeking despite negative consequences in addiction and repetitive binge eating phenomena in binge eating disorder (BED), both characterized by the pathological use of drugs or food. We further examine repetitive stereotyped actions to avoid negative outcomes in obsessive-compulsive disorder (OCD). To illustrate, such disorders may be characterized by a stimulus (e.g., the feel of a cigarette, dirty hands) associated with a specific response (e.g., smoking, hand washing). The behavior is initially goal directed, driven by the value of the outcome, which may be rewarding (e.g., the nicotine hit, clean hands), or to avoid an aversive outcome (e.g., anxiety of the obsession or urge). The pathological behavior represents a shift away from goal-directed control toward implicit habitual responses based on stimulus–response associations autonomous of the outcome values, including associated long-term negative outcomes (e.g., health consequences, missing work).

A compelling theory in addictions posits that drug consumption is initially goal directed, guided by drug-associated positive effects, and with chronic drug exposure, behavior becomes inflexible and habitual (24,25). Across multiple substances, including alcohol (26–28), cocaine (29,30), and amphetamines (31), exposure in rodents is associated with a shift from goal-directed to habitual behaviors paralleled by a shift from ventral to dorsal striatal engagement.

Human studies in alcohol misuse demonstrate effects on this shift in strategy dependent on the severity and recency of alcohol use and the duration of abstinence. Alcohol severity in a large sample of social drinkers was not associated with impaired model-based or model-free control, although an earlier drinking age was associated with enhanced putaminal reward prediction error activity underlying habit control (32). In a large online sample, which included subjects with a range of severity and varying recency of use, self-reported severity of alcohol use was associated with impairments in model-based control (33). Similarly, young severe binge drinkers were shown on the two-step task to shift toward model-free habit control. Crucially, subjects tested within 4 days of the last binge episode had significantly higher model-free scores and lower model-based scores relative to those tested more than 4 days after the last binge episode (34).

Model-based control appears to be modulated by abstinence and may play a predictive role in treatment outcome.

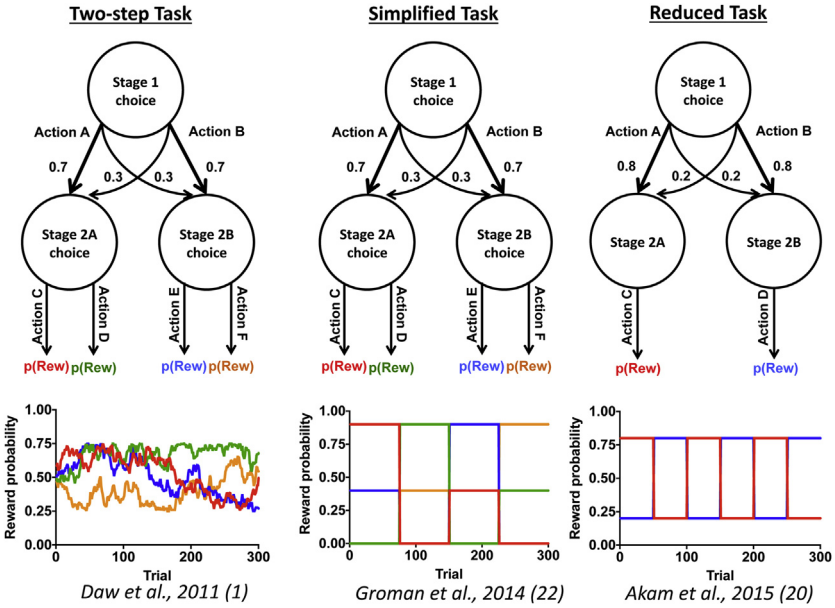

Figure 4. Two-step task in rodent models. The two-step task originally developed in human subjects has since been back-translated to animals (20–23). Although most of these translationally analogous tasks have yet to be published [but see (20)], a number of conference proceedings describe different task variants that have been successfully implemented in rats and mice. Versions of the two-step task that have been implemented in humans (left panel), rats (middle panel), and mice (right panel). In the two-stage task, subjects are presented with two stage 1 options, and responses on either of those options can lead to either a common or rare transition to stage 2. The rodent tasks use a similar structure, using either identical (22) or similar (20,21) translational probabilities. In stage 2, two variants of the task are available—rodents choose either between two options (22) or from a single option (20,21)—and choices in this stage are probabilistically reinforced with either food (22) or liquid (21) reward. These rodent variants share many of the critical features of the original human two-step task but with slight modifications that accommodate animal behavior. In humans, the payoff associated with the different stage 2 options are modified slowly and independently according to Gaussian random walks (1) to encourage learning and exploration. Animals, however, are less tolerant to stochastic schedules of reinforcement, particularly when there is only a minor benefit for optimal and cognitively taxing decisions (20). To increase the benefit of optimal decision making, the rodent two-stage tasks use an alternating schedule of reinforcement where the magnitude of the contrast between good and bad options is much greater than that in the human task. These rodent variants share many of the critical features of the original human two-step task but with slight modifications that accommodate animal behavior. p(Rew), probability by which reward is delivered.

Using the overtrained conflict task, alcohol-dependent (AD) subjects tested 24 hours abstinent showed impaired learning across trials requiring goal-directed as well as enhanced habitual behavior, suggesting a general impairment in response–outcome learning (35). The AD subjects also showed decreased vmPFC and anterior putaminal activity during goal-directed learning and showed increased posterior putaminal activity during habit learning (35). In one study, AD subjects tested 2 weeks abstinent revealed impaired model-based control (36), although these findings were not shown in two other studies with larger sample sizes (17,37). However, abstinence duration has been positively associated with model-based control (17). Furthermore, in AD subjects, subsequent relapse was predicted by an interaction between low model-based control and high alcohol expectancies, reflecting subjective expectations of the reinforcing effects of alcohol (37). Lower vmPFC activity to model-based state prediction error also predicted subsequent relapse in AD subjects.

Subjects with mixed substance dependence also show a shift from goal-directed to habitual control with overtraining (38). Actively using cocaine-dependent subjects, like actively using AD subjects, showed impairments in both goal-directed learning and enhanced habit learning to rewards on the slips-of-action task, with no differences on avoidance learning (39). Similarly, abstinent methamphetamine-dependent subjects also showed impaired model-based goal-directed control (17).

Taken together, these findings suggest that these processes are not impaired in social drinkers. Heavy alcohol exposure in alcohol misusers appears to impair goal-directed control and to enhance habit control, likely playing a role in maintenance of the addiction. The impairments are sensitive to abstinence and play a role in predicting treatment outcome. Behavioral measures of impaired model-based control interacting with high alcohol expectancies and vmPFC correlates of model-based control predict treatment outcome. Stimulant users show similar impairments in goal-directed and habit control persisting into abstinence.

BED is a recently accepted diagnosis in the DSM-5, defined by the rapid intake of a large amount of food with loss of control over the behavior and associated with negative consequences. Rodent binge-like models of food consumption fostered habit formation in a devaluation task (40). Similarly, obese humans with BED relative to those without BED had impaired model-based control (17) and lower gray matter volumes in regions implicated in model-based control (mOFC and caudate and lateral PFCs) (17). However, the specificity for binge eating is not completely clear given that obese subjects without BED have also shown impaired goal-directed control tested using a devaluation design (41). Self-reported pathological eating symptoms have also been associated with impaired model-based control in a large online sample (33).

Subjects with OCD also showed decreased sensitivity toward outcome devaluation in the slip-of-action task, indicating an overreliance on habits (42). Similarly, on the two-step task to reward outcomes, subjects with OCD were impaired in model-based control, a finding replicated at two sites, with compulsivity severity correlating negatively with model-based control and correlating positively with model-free control (17,43). Severity of self-reported obsessive-compulsive symptoms in a large online sample was also negatively associated with model-based behavior to reward (33). The

phenomenology of compulsive symptoms in OCD may also be captured by aversive avoidance habits, which may be of particular relevance. Using an aversive shock habit task, subjects with OCD showed greater habitual responding following overtraining to a virtual devaluation of a shock outcome (44). These results contrast with a study using a loss version of the two-step task where subjects with OCD demonstrated the opposite, that is, more pronounced model-based behaviors to avoid loss outcomes (43). These different findings may reflect task differences related to losses as compared with shock outcomes or habitual behaviors during early learning versus overtraining.

Building a model of the environment is a prerequisite for model-based control. This requires learning about environmental contingencies or the relationship of choice values and the integration of different choice options and their fictive consequences. Preliminary evidence suggests that the capacity to build an internal model of the environment might be impaired in addictive and other compulsive disorders; in rodents, cocaine self-administration abolishes the representation of inferred outcomes (45). In human studies, smokers do not guide their decisions by fictive prediction errors (46), and AD subjects do not update "what might have happened" particularly after punishment (47). Medial prefrontal blood oxygen level–dependent activation, reflecting prediction errors incorporating the updating of alternative choice options, was reduced in both AD subjects and those with BED (47,48). Similarly, the use of counterfactual computations in decision making was also diminished insubjects with OCD (49). These building blocks of an internal environmental model, and hence of model-based behavior, have been suggested as promising treatment targets (50).

These converging findings suggest that reduced goal-directed behavior may be a transdiagnostic feature across compulsive disorders. Recent studies suggest that other psychiatric disorders such as schizophrenia and social phobia, which are not commonly characterized by compulsions, also have impaired model-based control (51,52). These findings have challenged the specificity of the association between compulsivity and such impairments in strategy. However, a recent study using a factor analytic approach addressed this question of specificity in a large online dataset showing that reduced model-based control was indeed selectively associated with a transdiagnostic symptom dimension comprising compulsive behavior and intrusive thought (33).

NEUROCHEMICAL SUBSTRATES OF MULTISTEP TASKS

In rodent studies, dopamine enhances habit formation (31) related to dopamine D_1 receptor activation (53). Selective nigrostriatal dopaminergic lesions impair habit formation (54). However, in humans, a different picture emerges. Depleting the dopamine precursor with tyrosine depletion increases habitual control in the slips-of-action task (55). In Parkinson's disease, characterized by greater dopaminergic depletion of dorsal relative to ventral striatal regions, impairments in habit control have not been demonstrated (56,57). Instead, the severity of Parkinson's disease is associated with impairments in goal-directed control (56) and patients tested off medications show impaired model-based control that improves with dopaminergic medications (57). Healthy humans challenged with levodopa also show enhanced model-based control (58). Similarly, greater ventral striatal presynaptic dopamine synthesis, measured using ^{18}F-dihydroxyphenylalanine positron emission tomography, correlates with greater model-based control (59). These human studies contrast with the preclinical literature and may reflect several plausible mechanisms (55,57). Dopaminergic medication challenges in humans have limited anatomical specificity and may act on regions implicated in model-based control. Neural regions subserving model-based and model-free control also overlap. Human tasks are also relatively undertrained compared with rodent tasks.

A role for serotonin (5-hydroxytryptamine [5-HT]) in habit control has also been reported. Decreasing forebrain 5-HT and systemic 5-HT_{2C} antagonism enhanced compulsive cocaine seeking in rodents, which was reversed by a 5-HT_{2C} agonist (60). Greater compulsive cocaine-seeking behaviors were also reversible with citalopram, a selective serotonin reuptake inhibitor. Overexpression of 5-HT_6 receptors in the rodent dorsolateral striatum was associated with decreased habit control (61). Similarly, in healthy humans, central serotonin depletion with tryptophan depletion enhances habitual responding on the slips-of-action task (62). Tryptophan depletion also impaired model-based control to reward outcomes but enhanced model-free control to loss outcomes in the two-step task. One possible mechanism whereby tonic 5-HT might enhance goal-directed behaviors is by changing the long-run average reward representation to provide a positive or negative signal of the "goodness" or "badness" of the environment (63). 5-HT signaling may also signify the cost associated with deliberation (64,65).

CHARACTERISTICS OF MULTISTEP TASKS

Several lines of evidence suggest that the model-based measure in the two-step task appropriately captures goal-directed control. Support for commonalities between the model-free measure and conventional overtraining and devaluation tasks exists but is more mixed.

The two-step task in healthy control subjects correlates with a conventional overtraining and food outcome devaluation task, which most closely approximates rodent models (66). However, factors commonly known to influence the shift from goal-directed to habitual behaviors appear to have a greater effect on the model-based measure but not on the model-free measure. Thus, an acute stress challenge impaired model-based control in healthy control subjects, an effect mediated by working memory and stress (67,68). Model-based control also depends on individual cognitive capacity (e.g., processing speed, dual task performance) (69,70). Outcome manipulations also appear to preferentially affect the model-based system. Model-based control was shown to predict sensitivity to outcome devaluation (71). Similarly, greater outcome salience (e.g., greater reward value or losses relative to gains) enhanced model-based control (43).

A developmental trajectory in which model-free control on the two-step task is present across all neurodevelopmental stages, but model-based control is absent in children, emerges

in adolescents and increases in adults (72). This may also reflect the underlying development of working memory and cognitive control.

Taken together, the model-based measure of the two-step task appears to appropriately formalize the measure of goal-directed control and has greater capacity for change. In contrast, the model-free measure describes actions based on previously reinforced learning during the early stages of learning, but its relationship to conventional overtraining and devaluation habit procedures is less well developed.

The model-based measure is also not without its limitations. Model-based decision making does not increase accuracy in the two-step task, suggesting that the computational cost to use the model-based strategy does not necessarily pay off (73). Furthermore, the model-based system uses explicit knowledge of the task structure and transition matrix. The capacity to learn model-based representations is not assessed in commonly used versions of the two-step task. The test–retest reliability of the task also has not been clearly established.

PAVLOVIAN LEARNING: MODEL-BASED CONTROL AND THE ORBITOFRONTAL CORTEX

Beyond instrumental learning, the broader construct of inferred model-based control (reflecting imagined value based on knowledge of the associative task structure) versus model-free control (relying on stored values) is also relevant to pavlovian processes and is shown to be dissociable in rodent studies (45,74,75) (Figure 5). These processes may be particularly relevant for addictions in which drug-related cues may represent learned pavlovian associations between neutral and unconditioned stimuli without instrumental responding. The learned cue (e.g., cigarette smoke), known as the conditioned stimulus, paired with the unconditioned stimulus (e.g., nicotine reward) triggers conditioned responses (e.g., autonomic activity) that might lead to urges, craving, and relapses. Model-based pavlovian learning, similar to structural studies of the instrumental two-step task, is associated with the rodent OFC and hypothesized to reflect the capacity to integrate outcome expectancies with the associative task structure (45,74,75). Rodents trained to self-administer cocaine also show impairments in inferred model-based learning on behavioral (76) and neurophysiological measures of model-based pavlovian tasks (45).

CONCLUSIONS

Model-based and model-free control in instrumental (goal-directed and habit learning) and pavlovian learning processes shows translational clinical relevance for dimensional psychiatry. The model-based measure in sequential decision tasks appears to formalize the measure of goal-directed control and is impaired across compulsive behaviors. Habit learning from overtrained and devaluation habit procedures have similarly shown enhancement in compulsive behaviors. The overlapping construct of model-free control, which describes actions based on previously reinforced learning during the early stages of learning, shows some impairments in psychiatric disorders, but more work is required.

Dissociable frontostriatal regions associated with goal-directed and habitual behaviors overlap between rodent and human studies. Model-based control extends these neural processes to higher order associative regions, and engagement of the OFC highlights potential overlaps with model-based pavlovian learning. The integration of strategies

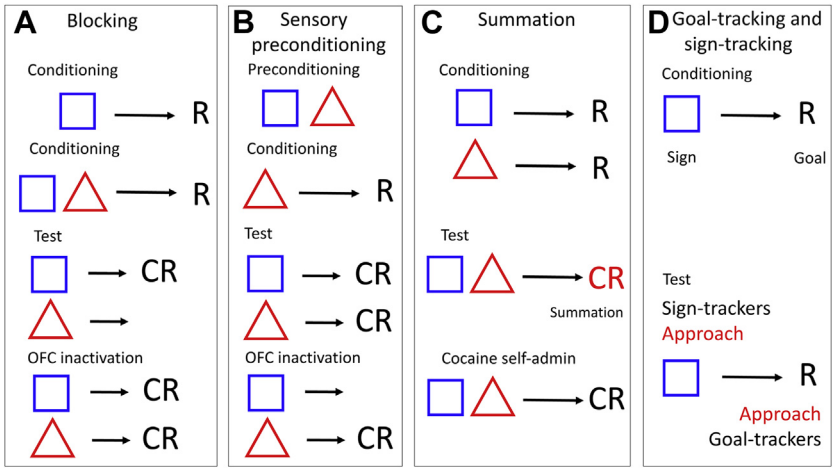

Figure 5. Model-based pavlovian learning and the orbitofrontal cortex. Model-based learning can be dissociated from stored values during pavlovian learning. The following tasks test the inference of values based on the knowledge of the associative model of the task rather than stored knowledge. Rodents trained to self-administer cocaine show the same impairments as orbitofrontal cortex (OFC) inactivation, suggesting impaired integration of outcome expectancies requiring inference (76). **(A)** Blocking task. A cue is first taught to predict reward (R) and later is presented with a new cue followed by reward. During testing, the new cue exhibits little conditioned responding (CR) as the original cue predicting reward blocks learning. OFC inactivation increases responding to both the control cue and the blocked cue. Similarly, using an unblocking paradigm in which value (differing number of pellets) and identity (different flavors) are differentiated, the ventral striatum is necessary for learning changes in reward identity and value, and the OFC is necessary for learning changes in reward identity (74). **(B)** Sensory preconditioning. A rodent learns to associate two sensory cues, one of which is later associated with food reward. The CR to the preconditioned cue reflects inferred value based on knowledge of the associative task structure, whereas the reward-paired cue reflects stored knowledge. OFC inactivation impairs the CR to the preconditioned cue but leaves intact response to the reward-paired cue (75). **(C)** Pavlovian overexpectation tasks. Cues are first conditioned to predict reward followed by compound cue training in which two cues are presented together followed by the same reward. The compound cue elicits enhanced responding, also known as summation, in which a novel expectation for increased reward is inferred and has not yet been experienced. In rodents trained to self-administer cocaine relative to sucrose, single unit activity and pyramidal neuron excitability in the OFC is impaired to the compound cue (45). **(D)** Goal-directed and sign-tracking rodents. The dissociation of goal tracking and sign tracking has also been associated with model-based learning (79).

converging on the ventral striatum and vmPFC is particularly relevant when the trade-off between strategies is emphasized. Back-translation of the two-step task in rodents has reasonable fidelity with the human analogue and holds intriguing potential.

The role of limited cognitive resources, such as working memory or the tendency to use the least effortful strategy, is relevant particularly for patient studies. Further studies in patient populations focusing on the relative balance of reward and aversive outcomes or other salient outcomes and the role of stress, working memory, and cognitive control in the shift between strategies are indicated. Thus, treatment targets can include the upstream cause of the strategy shift, sensitivity to the outcome valence, or underlying core impairments in specific strategies. Focusing on shifting model-free control would be therapeutically relevant. Further optimization of sequential learning tasks is required to understand the relationship with cognitive control, to simplify the task for patient testing, to optimize the cost–benefit ratio for model-based control, and to capture a measure that more closely represents conventional habit definitions. The relationship between model-based control in instrumental learning and that in pavlovian learning remains to be further elucidated.

More broadly, the capacity to build an internal model of the task structure or environment can potentially be extended to other forms of behavioral inflexibility such as set shifting and reversal learning and forms of impulsivity such as reflection impulsivity (77). Taken together, these findings highlight a role for model-based control as a transdiagnostic impairment relevant to dimensional psychiatry and represent a promising therapeutic target.

ACKNOWLEDGMENTS AND DISCLOSURES

VV is funded by a Medical Research Council Senior Clinical Fellowship. MS and AR are supported by the German Research Foundation (Deutsche Forschungsgemeinschaft).

VV has received fees as an expert court witness for proceedings related to a dopamine agonist. AR, MS, and SG report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Department of Psychiatry (VV), University of Cambridge, and Cambridgeshire and Peterborough NHS Foundation Trust (VV), Cambridge, United Kingdom; Lifespan Developmental Neuroscience (AR), Department of Psychology, Dresden, Department of Neurology (AR), Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, and Department of Psychiatry and Psychotherapy (MS), Charite–Universitatsmedizin Berlin, Berlin, Germany; and Department of Psychiatry (SG), Yale University, New Haven, Connecticut.

Address correspondence to Valerie Voon, M.D., Ph.D., F.R.C.P.C., Department of Psychiatry, University of Cambridge, Addenbrookes Hospital, Level E4, Box 189, Hills Road, Cambridge CB2 0QQ, UK; E-mail: voonval@gmail.com.

Received Dec 14, 2016; revised Apr 11, 2017; accepted Apr 12, 2017.

Supplementary material cited in this article is available online at http://dx.doi.org/10.1016/j.biopsych.2017.04.006.

REFERENCES

1. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011): Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204–1215.
2. Glascher J, Daw N, Dayan P, O'Doherty JP (2010): States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585–595.
3. Lucantonio F, Stalnaker TA, Shaham Y, Niv Y, Schoenbaum G (2012): The impact of orbitofrontal dysfunction on cocaine addiction. Nat Neurosci 15:358–366.
4. Adams CD, Dickinson A (1981): Instrumental responding following reinforcer devaluation. Q J Exp Psychol 33:109–122.
5. Dickinson A, Balleine BW (2002): The role of learning in the operation of motivational systems. In: Gallister CR, editor. Stevens' Handbook of Experimental Psychology: Learning, Motivation and Emotion, 3rd ed. New York: John Wiley, 497–534.
6. de Wit S, Corlett PR, Aitken MR, Dickinson A, Fletcher PC (2009): Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. J Neurosci 29:11330–11338.
7. Dolan RJ, Dayan P (2013): Goals and habits in the brain. Neuron 80:312–325.
8. Yin HH, Ostlund SB, Knowlton BJ, Balleine BW (2005): The role of the dorsomedial striatum in instrumental conditioning. Eur J Neurosci 22:513–523.
9. Balleine BW, Dickinson A (1998): Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. Neuropharmacology 37:407–419.
10. Yin HH, Knowlton BJ, Balleine BW (2004): Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur J Neurosci 19:181–189.
11. Killcross S, Coutureau E (2003): Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb Cortex 13:400–408.
12. Tricomi E, Balleine BW, O'Doherty JP (2009): A specific role for posterior dorsolateral striatum in human habit learning. Eur J Neurosci 29:2225–2232.
13. Horga G, Maia TV, Marsh R, Hao X, Xu D, Duan Y, *et al.* (2015): Changes in corticostriatal connectivity during reinforcement learning in humans. Hum Brain Mapp 36:793–803.
14. de Wit S, Watson P, Harsay HA, Cohen MX, van de Vijver I, Ridderinkhof KR (2012): Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. J Neurosci 32:12066–12075.
15. Smittenaar P, FitzGerald TH, Romei V, Wright ND, Dolan RJ (2013): Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. Neuron 80:914–919.
16. Wunderlich K, Dayan P, Dolan RJ (2012): Mapping value based planning and extensively trained choice in the human brain. Nat Neurosci 15:786–791.
17. Voon V, Derbyshire K, Ruck C, Irvine MA, Worbe Y, Enander J, *et al.* (2015): Disorders of compulsivity: A common bias towards learning habits. Mol Psychiatry 20:345–352.
18. Morris LS, Kundu P, Dowell N, Mechelmans DJ, Favre P, Irvine MA, *et al.* (2016): Fronto-striatal organization: Defining functional and microstructural substrates of behavioural flexibility. Cortex 74:118–133.
19. Valentin VV, Dickinson A, O'Doherty JP (2007): Determining the neural substrates of goal-directed learning in the human brain. J Neurosci 27:4019–4026.
20. Akam T, Costa R, Dayan P (2015): Simple plans or sophisticated habits? State, transition and learning interactions in the two-step task. PLoS Comput Biol 11:e1004648.
21. Miller K, Akrami A, Botvinick M, Brody C (2015): The role of the orbitofrontal cortex in model-based planning in the rat. Presented at the annual meeting of the Society for Neuroscience, October 17–21, Chicago.
22. Groman SM, Chen L, Smith NJ, Lee D, Taylor JR (2014): Dorsomedial striatum lesions disrupt the balance between model-free and model-based learning in a multi-stage decision-making task. Presented at the annual meeting of the Society for Neuroscience, November 15–19, Washington, DC.
23. Miranda B, Malalasekera N, Dayan P, Kennerley S (2014): Evidence of model-based and model-free reinforcement learning in prefrontal cortex and striatal neurons. Presented at the annual meeting of the Society for Neuroscience, November 15–19, Washington, DC.

24. Everitt BJ, Robbins TW (2005): Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. Nat Neurosci 8:1481–1489.
25. Dayan P (2009): Dopamine, reinforcement learning, and addiction. Pharmacopsychiatry 42(suppl 1):S56–S65.
26. Dickinson A, Wood N, Smith JW (2002): Alcohol seeking by rats: Action or habit? Q J Exp Psychol B 55:331–348.
27. Lopez MF, Becker HC, Chandler LJ (2014): Repeated episodes of chronic intermittent ethanol promote insensitivity to devaluation of the reinforcing effect of ethanol. Alcohol 48:639–645.
28. Corbit LH, Nie H, Janak PH (2012): Habitual alcohol seeking: Time course and the contribution of subregions of the dorsal striatum. Biol Psychiatry 72:389–395.
29. Zapata A, Minney VL, Shippenberg TS (2010): Shift from goal-directed to habitual cocaine seeking after prolonged experience in rats. J Neurosci 30:15457–15463.
30. Schmitzer-Torbert N, Apostolidis S, Amoa R, O'Rear C, Kaster M, Stowers J, *et al.* (2015): Post-training cocaine administration facilitates habit learning and requires the infralimbic cortex and dorsolateral striatum. Neurobiol Learn Mem 118:105–112.
31. Nelson A, Killcross S (2006): Amphetamine exposure enhances habit formation. J Neurosci 26:3805–3812.
32. Nebe S, Kroemer N, Schad D, Bernhardt N, Sebold M, Muller D, *et al.* (2017): No association of goal-directed and habitual control with alcohol consumption in young adults [published online ahead of print Jan 23]. Addict Biol.
33. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND (2016): Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. eLife 5:e11305.
34. Doñamayor N, Strelchuk D, Baek K, Banca P, Voon V (2017): The involuntary nature of binge drinking: Goal directedness and awareness of intention [published online ahead of print Apr 16]. Addict Biol.
35. Sjoerds Z, de Wit S, van den Brink W, Robbins TW, Beekman AT, Penninx BW, *et al.* (2013): Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. Transl Psychiatry 3:e337.
36. Sebold M, Deserno L, Nebe S, Schad DJ, Garbusow M, Hagele C, *et al.* (2014): Model-based and model-free decisions in alcohol dependence. Neuropsychobiology 70:122–131.
37. Sebold M, Nebe S, Garbusow M, Guggenmos M, Schad D, Beck A, Kuitunen-Paul S, *et al.* (2017): When habits are dangerous: Alcohol expectancies and habitual decision making predict relapse in alcohol dependence [published online ahead of print May 22]. Biol Psychiatry.
38. McKim TH, Bauer DJ, Boettiger CA (2016): Addiction history associates with the propensity to form habits. J Cogn Neurosci 28:1024–1038.
39. Ersche KD, Gillan CM, Jones PS, Williams GB, Ward LH, Luijten M, *et al.* (2016): Carrots and sticks fail to change behavior in cocaine addiction. Science 352:1468–1471.
40. Furlong TM, Jayaweera HK, Balleine BW, Corbit LH (2014): Binge-like consumption of a palatable food accelerates habitual control of behavior and is dependent on activation of the dorsolateral striatum. J Neurosci 34:5012–5022.
41. Horstmann A, Dietrich A, Mathar D, Possel M, Villringer A, Neumann J (2015): Slave to habit? Obesity is associated with decreased behavioural sensitivity to reward devaluation. Appetite 87:175–183.
42. Gillan CM, Papmeyer M, Morein-Zamir S, Sahakian BJ, Fineberg NA, Robbins TW, *et al.* (2011): Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. Am J Psychiatry 168:718–726.
43. Voon V, Baek K, Enander J, Worbe Y, Morris LS, Harrison NA, *et al.* (2015): Motivation and value influences in the relative balance of goal-directed and habitual behaviours in obsessive-compulsive disorder. Transl Psychiatry 5:e670.
44. Gillan CM, Morein-Zamir S, Urcelay GP, Sule A, Voon V, Apergis-Schoute AM, *et al.* (2014): Enhanced avoidance habits in obsessive-compulsive disorder. Biol Psychiatry 75:631–638.
45. Lucantonio F, Takahashi YK, Hoffman AF, Chang CY, Bali-Chaudhary S, Shaham Y, *et al.* (2014): Orbitofrontal activation restores insight lost after cocaine use. Nat Neurosci 17:1092–1099.
46. Chiu PH, Lohrenz TM, Montague PR (2008): Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. Nat Neurosci 11:514–520.
47. Reiter AMF, Deserno L, Kallert T, Heinz A, Heinze HJ, Schlagenhauf F (2016): Behavioral and neural signatures of reduced updating of alternative options in alcohol-dependent patients during flexible decision-making. J Neurosci 36:10935–10948.
48. Reiter A, Heinze H-J, Schlagenhauf F, Deserno L (2017): Impaired flexible reward-based decision-making in binge eating disorder: Evidence from computational modeling and functional neuroimaging. Neuropsychopharmacology 42:628–637.
49. Gillan CM, Morein-Zamir S, Kaser M, Fineberg NA, Sule A, Sahakian BJ, *et al.* (2014): Counterfactual processing of economic action–outcome alternatives in obsessive-compulsive disorder: Further evidence of impaired goal-directed behavior. Biol Psychiatry 75:639–646.
50. Schoenbaum G, Chang C-Y, Lucantonio F, Takahashi YK (2016): Thinking outside the box: Orbitofrontal cortex, imagination, and how we can treat addiction. Neuropsychopharmacology 41: 2966–2976.
51. Alvares GA, Balleine BW, Guastella AJ (2014): Impairments in goal-directed actions predict treatment response to cognitive-behavioral therapy in social anxiety disorder. PLoS One 9:e94778.
52. Culbreth AJ, Westbrook A, Daw ND, Botvinick M, Barch DM (2016): Reduced model-based decision-making in schizophrenia. J Abnorm Psychol 125:777–787.
53. Nelson AJ, Killcross S (2013): Accelerated habit formation following amphetamine exposure is reversed by D1, but enhanced by D2, receptor antagonists. Front Neurosci 7:76.
54. Faure A, Haberland U, Conde F, El Massioui N (2005): Lesion to the nigrostriatal dopamine system disrupts stimulus–response habit formation. J Neurosci 25:2771–2780.
55. de Wit S, Standing HR, Devito EE, Robinson OJ, Ridderinkhof KR, Robbins TW, *et al.* (2012): Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. Psychopharmacology (Berl) 219:621–631.
56. de Wit S, Barker RA, Dickinson T, Cools R (2011): Habitual versus goal-directed action control in Parkinson's disease. J Cogn Neurosci 23:1218–1229.
57. Sharp ME, Foerde K, Daw ND, Shohamy D (2016): Dopamine selectively remediates "model-based" reward learning: A computational approach. Brain 139:355–364.
58. Wunderlich K, Smittenaar P, Dolan RJ (2012): Dopamine enhances model-based over model-free choice behavior. Neuron 75: 418–424.
59. Deserno L, Huys QJ, Boehme R, Buchert R, Heinze HJ, Grace AA, *et al.* (2015): Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. Proc Natl Acad Sci U S A 112:1595–1600.
60. Pelloux Y, Dilleen R, Economidou D, Theobald D, Everitt BJ (2012): Reduced forebrain serotonin transmission is causally involved in the development of compulsive cocaine seeking in rats. Neuropsychopharmacology 37:2505–2514.
61. Eskenazi D, Neumaier JF (2011): Increased expression of 5-HT_6 receptors in dorsolateral striatum decreases habitual lever pressing, but does not affect learning acquisition of simple operant tasks in rats. Eur J Neurosci 34:343–351.
62. Worbe Y, Savulich G, de Wit S, Fernandez-Egea E, Robbins TW (2015): Tryptophan depletion promotes habitual over goal-directed control of appetitive responding in humans. Int J Neuropsychopharmacol 18:pyv013.
63. Daw ND, Kakade S, Dayan P (2002): Opponent interactions between serotonin and dopamine. Neural Netw 15:603–616.
64. Niv Y, Daw ND, Joel D, Dayan P (2007): Tonic dopamine: Opportunity costs and the control of response vigor. Psychopharmacology (Berl) 191:507–520.
65. Keramati M, Dezfouli A, Piray P (2011): Speed/accuracy trade-off between the habitual and the goal-directed processes. PLoS Comput Biol 7:e1002055.

66. Friedel E, Koch SP, Wendt J, Heinz A, Deserno L, Schlagenhauf F (2014): Devaluation and sequential decisions: Linking goal-directed and model-based behavior. Front Hum Neurosci 8:587.
67. Otto AR, Raio CM, Chiang A, Phelps EA, Daw ND (2013): Working-memory capacity protects model-based learning from stress. Proc Natl Acad Sci U S A 110:20941–20946.
68. Radenbach C, Reiter AM, Engert V, Sjoerds Z, Villringer A, Heinze HJ, *et al.* (2015): The interaction of acute and chronic stress impairs model-based behavioral control. Psychoneuroendocrinology 53:268–280.
69. Otto AR, Gershman SJ, Markman AB, Daw ND (2013): The curse of planning: Dissecting multiple reinforcement-learning systems by taxing the central executive. Psychol Sci 24:751–761.
70. Schad DJ, Junger E, Sebold M, Garbusow M, Bernhardt N, Javadi AH, *et al.* (2014): Processing speed enhances model-based over model-free reinforcement learning in the presence of high working memory functioning. Front Psychol 5:1450.
71. Gillan CM, Otto AR, Phelps EA, Daw ND (2015): Model-based learning protects against forming habits. Cogn Affect Behav Neurosci 15: 523–536.
72. Decker JH, Otto AR, Daw ND, Hartley CA (2016): From creatures of habit to goal-directed learners: Tracking the developmental emergence of model-based reinforcement learning. Psychol Sci 27:848–858.
73. Kool W, Cushman FA, Gershman SJ (2016): When does model-based control pay off? PLoS Comput Biol 12:e1005090.
74. McDannald MA, Takahashi YK, Lopatina N, Pietras BW, Jones JL, Schoenbaum G (2012): Model-based learning and the contribution of the orbitofrontal cortex to the model-free world. Eur J Neurosci 35:991–996.
75. Jones JL, Esber GR, McDannald MA, Gruber AJ, Hernandez A, Mirenzi A, *et al.* (2012): Orbitofrontal cortex supports behavior and learning using inferred but not cached values. Science 338:953–956.
76. Wied HM, Jones JL, Cooch NK, Berg BA, Schoenbaum G (2013): Disruption of model-based behavior and learning by cocaine self-administration in rats. Psychopharmacology (Berl) 229:493–501.
77. Banca P, Lange I, Worbe Y, Howell NA, Irvine M, Harrison NA, *et al.* (2016): Reflection impulsivity in binge drinking: Behavioural and volumetric correlates. Addict Biol 21:504–515.
78. Blechert J, Meule A, Busch NA, Ohla K (2014): Food-pics: An image database for experimental research on eating and appetite. Front Psychol 5:617.
79. Lesaint F, Sigaud O, Flagel SB, Robinson TE, Khamassi M (2014): Modelling individual differences in the form of Pavlovian conditioned approach responses: A dual learning systems approach with factored representations. PLoS Comput Biol 10:e1003466.