Neurobiology and computational structure of decision-making in addiction

Quentin JM Huys[1,2,3,4], Anne Beck[5], Peter Dayan[3] and Andreas Heinz[5]

1 Translational Neuromodeling Unit, ETH Zurich and University of Zurich, Switzerland
2 Department of Psychiatry, Psychotherapy and Psychosomatics, Hospital of Psychiatry, University of Zurich, Switzerland
3 Gatsby Computational Neuroscience Unit, University College London, UK
4 Wellcome Trust Centre for Neuroimaging, University College London, UK
5 Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin Berlin, Germany

Abstract

An increasing wealth of experimental detail is becoming available about the development and nature of addiction. Critical issues such as the varying vulnerabilities of individuals who develop addiction are being illuminated across levels of phenomenological, psychological and neurobiological detail. Furthermore, a rich theoretical understanding is emerging in the field of neural reinforcement learning, with glimmers as to how this might be related to the subjective experience of those individuals affected. In this chapter, we consider some particularly pressing current issues in the interface between experiment and theory, notably the so-called "compulsive" phase of drug taking.

Keywords: Addiction, Dopamine, Computational Psychiatry, Reinforcement learning, cached values, D2 receptors, Sign tracking.

Introduction

The development of addiction is characterized by several key features (Beck et al., 2011; Gelder et al., 2006; Heinz et al., 2009a; Koob, 2003; DSM IV-TR (American Psychiatric Association, 1994); ICD-10 (World Health Organization, 1990)). These include an evolving tolerance to the effects of the drug of abuse, which is accompanied by adaptations in central neurotransmission. Strikingly, tolerance does not decrease, but rather seems to coincide with an increase in the lure of drugs. Once established, addictions are accompanied by the subjective experience of intense cravings for the drug and by elaborate drug seeking behaviors. By comparison with drugs, normal reinforcers lose the ability to control behavior: people with substance dependencies continue to take drugs despite apparently understanding and acknowledging the devastating effect this may have on their life and despite expressed desires and frequent attempts to abstain. Indeed, "wanting" and "liking" the drug appear to be separable phenomena: addicts may want drugs even when they no longer expect or experience positive hedonic effects (Robinson and Berridge, 2003). Thus, the objective behavior and the subjective experience of addictive disorders evidence a certain disconnect. Though largely preclinical, recent studies have resulted in an increasingly detailed description of how initially harmless drug-taking for hedonic reasons is transformed into a maladaptive and "compulsive" pattern of "wanting" to take drugs that is resistant to negative outcomes (Vanderschuren and Everitt, 2004; Everitt and Robbins, 2005). Finally, a key feature of addiction is the tendency for relapse years or decades into abstinence. This chapter will describe how models of decision-making can be used to bridge the gap between neurobiological mechanisms, behavior and possibly even individual subjective experience.

The main accounts of decision making come from the evolving field of neural reinforcement learning (Sutton and Barto, 1998), which links computational notions of optimal control (Puterman, 2005) and psychological data on conditioning (Sutton and Barto, 1990), together with their neural substrate in the striatum and the latter's associated neuromodulatory, amygdala and cortical inputs (Niv, 2009; Daw and Doya, 2006). There is now extensive evidence that addictive substances modulate and derail adaptive decision-making in part by influencing the function of dopamine. For instance, drugs of abuse share a prominent ability to release dopamine (Koob, 1992), and electrical as well as optogenetic stimulation of dopamine release can lead to behavior that shares important characteristics with addictions (Olds, 1956; Tsai et al., 2009; though see Garris et al., 1999). Furthermore, as we review below, imaging has shown that there are abnormalities in dopamine D_2 receptors in humans (Heinz et al., 1996; Volkow et al., 1996) of a sort that various animal models show can contribute to addictive behaviors (Thanos et al., 2001; Morgan et al., 2002; Johnson and Kenny, 2010; Bello et al., 2011). The stages of drug addiction, from early in the addiction up through "compulsive" drug taking and to relapse, provide clues but also complications for neural accounts of reinforcement learning. Most explanations of the onset of drug consumption are straightforward, with drugs acting as strong surrogate rewards (Redish, 2004). However, what poses greater challenges (Redish et al., 2008; Dayan 2009) is the development of deeply ingrained, recidivist states and the involvement of dopamine therein.

In this chapter, we focus on the so-called "compulsive" nature of drug taking, i.e. the seeking and taking of drugs despite severely aversive consequences. We discuss recent conceptions of the interaction of different instrumental and Pavlovian learning systems, and the way that dopamine manipulations occasioned by drugs may affect both. These issues are certainly not settled – there are, for instance, debates concerning the migration of control from the ventral to dorsal striatum (Everitt et al., 2008) versus enhanced incentive salience of drug-associated cues (Robinson and Berridge, 2001, 2003) that have yet to be fully resolved.

At the outset, it is important to clarify the use of the word "compulsion" especially as it applies to the subjective experience of the patient. The clinical description of compulsions in the setting of obsessive compulsive disorder (OCD) denotes behaviors whose aims are to reduce negative affect associated with

aversive obsessive thoughts (See chapter *by Denys, Prosée and Stein,* this volume*)*. That is, in OCD, compulsions denote behaviors aimed at terminating or reducing aversive states. In addiction research, "compulsive" is used rather to denote behaviors persisting despite their aversive consequences (Robbins and Everitt, 2005; Robbins et al., 2012). They do not refer to behaviors that might be driven by aversive withdrawal states (Koob, 2003). These two different compulsions likely have very different neural counterparts, with OCD, for instance, being associated with increased - but drug addiction with decreased - frontal cortical resting state activation (Baxter et al., 1988; Heinz, 1999; Saxena et al., 1998; Volkow et al., 1993; 2011). What both descriptions do however capture is the profoundly compelling force to act. We will leave the similarity of compulsions in addiction and OCD for further consideration elsewhere. Here, we focus on the notion of compulsions in addiction because they are key to the development of learning-theoretic models of addiction, and because they provide one possible avenue for "wanting" and seeking drugs despite not "liking" them, that is for the disconnect between subjective experience and observable behavior.

Reinforcement learning

Reinforcement learning (RL) comprises a set of approaches to adaptive optimal control. These allow natural and artificial agents alike to learn to choose actions that maximize their rewards and minimize their punishments (Sutton and Barto, 1998). RL's closest psychological counterpart is thus instrumental conditioning (Mackintosh, 1983).

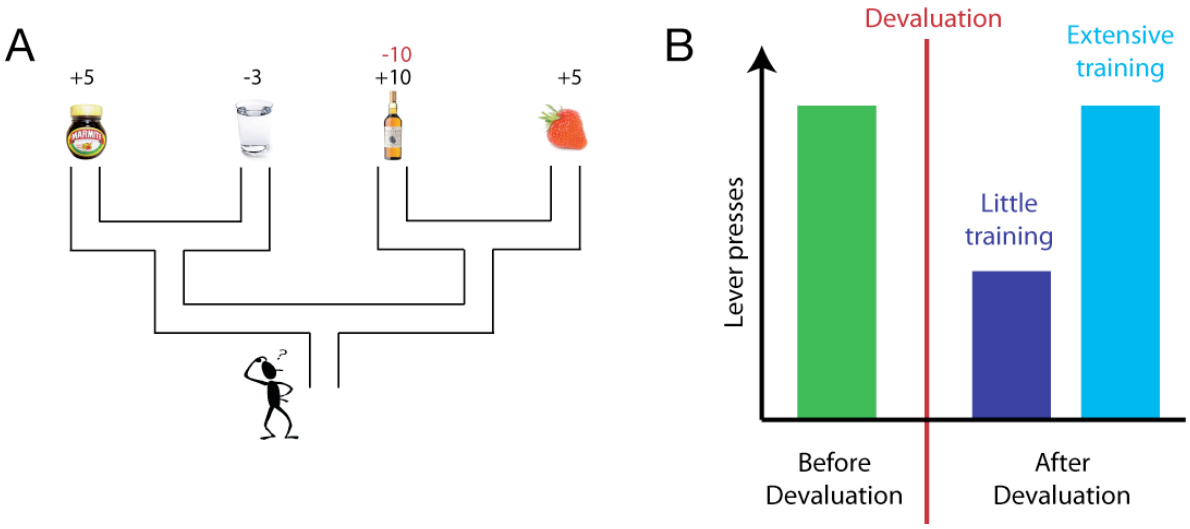

Figure 1: Model-based and model-free choices. A: Decision trees. Imagine having to navigate a maze. If one knew the layout of the maze, and knew which outcome was at the bottom of each path, one could consider all sequences of left-right choices to decide on the best one. For this maze, one would have to consider four sequences, each containing two choices. More generally, the tree could be much deeper, and there could be more options to choose from at each choice-point. In chess, for instance, the tree would not have 4, but on the order of 30^{40} branches. B: In outcome devaluation, rats first learn to press a lever to obtain a reward (green bar). The reward is then devalued, for instance through satiation or through association with illness. Rats are then given the option to press the lever again. If rats have been trained only a little, the rate at which they press the lever will be greatly reduced (dark blue bar). Responding is said to be goal-directed, here driven by the desirability of the outcome obtained by the action. If, on the other hand, the rats have been pressing the lever for a long time, they will not show a suppression of lever pressing, but continue pressing the lever for an outcome they no longer desire. Responding is now said to be habitual, and insensitive to the desirability of the outcome resulting from the action. Accruing experience thus results in a shift from an early, goal-directed, to a later, habitual phase.

RL focuses on the fact that most decision problems have longer-term, as well as immediate, consequences, and the benefits and costs of all of these have to be taken into account. This is difficult because the number of trajectories of future states and choices typically grows exponentially with the horizon. Consider chess: to deduce, purely from a knowledge of the rules, the best possible move, requires consideration of the future. If the player faces a choice of 30 possibilities at each move, then looking just two moves ahead requires consideration of 30 x 30 = 900 different combinations. Deducing the optimal move demands looking ahead all the way to the end of the game. If an average game is around 40 moves long, then around 30^{40} sequences have to be evaluated.

In so-called model-based decision-making, agents are assumed to have, or to be learning, a full description of an environment. Then, they explicitly consider all consequences of all future actions

starting from a state or location in the world, and use this to deduce which sequence of choices would be optimal. Figure 1a shows an example. In chess, a model-based strategy would correspond to knowing the rules and the aim of the game, and deducing from this alone the best possible action choice. Actions are thus taken with the explicit aim of achieving a particular outcome, given the known consequences of choices. This makes model-based RL exhibit the signature characteristic of goal-directed actions, namely instant sensitivity to changes in outcome values (green and dark blue bars in Figure 1b; Dickinson and Balleine, 2002; Daw et al., 2005; Dickinson, 1985; Valentin et al., 2007; de Wit et al., 2009; Tricomi et al., 2009). However, although model-based approaches are often guaranteed to produce good outcomes, they are typically rendered useless by their intensive computational demands, and thus cannot simply be solved by the model-based system. Indeed, animals and humans can be shown to employ model-based choice strategies, but only in rather small environments (Shallice, 1982; Simon and Daw, 2011, Huys et al. 2012).

At the other end of the spectrum of computational requirements lie model-free techniques. Put simply, these methods lead to repetition of what was rewarded in the past. To return to the chess example, experience of games in the past might teach the player that moves which result in the loss of the most powerful figure, the queen, are most likely to be deleterious, and should not be considered further. Model-free methods learn from past experience by maintaining a running average – a cache – of past rewards and punishments of each move at each state. They can adopt a particular manoeuvre called temporal difference learning (Sutton 1988) to do the updating and averaging efficiently. Indeed, given the assumption that the environment is fixed, particular ways of averaging over experiences can solve the hard problem of considering the future consequences of each action (Sutton and Barto, 1998). Consider again Figure 1a: In the absence of explicit knowledge about the maze, one could repeatedly walk through it, and, on each trial, iteratively update the estimate of the long-term outcomes of taking a left or right turn at each state. This would lead to an average estimate - through repeated experience - of the outcomes. Clearly, this would shift the onus from thinking through all possible outcomes to experiencing them multiple times.

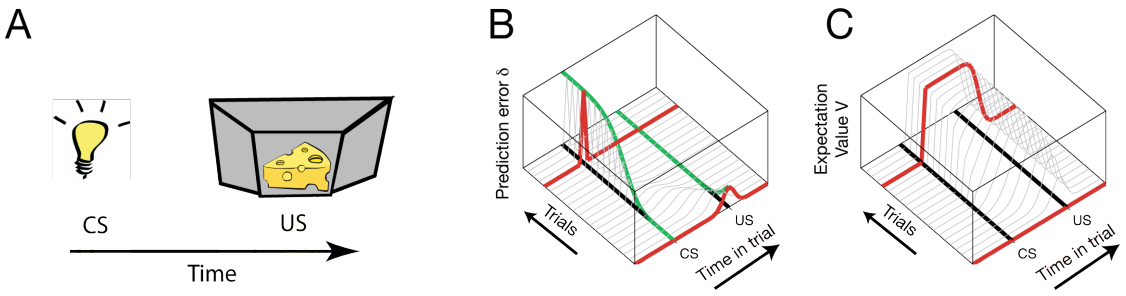

Figure 2: Temporal prediction error account of classical conditioning. A: A conditioned stimulus (CS) temporally precedes and predicts presentation of an unconditioned stimulus (a cheese reward delivered in a goal box for a mouse). B: Temporal prediction errors. During early trials, a prediction error arises at the time of the US (red line towards the right). As learning proceeds, this error signal moves towards the time of the CS onset (red line on the left). Critically, this shift of the prediction error signal is thus reflected, over trials, by a decay *of the error signal at the time of the US and an* increase *at the time of the CS (green lines). C: Time course of the cached expectations V_t over time. Initially, there is no expectation (flat red line to the right). After learning, the reward expectation rises on presentation of the CS and remains high until the US is received.*

Temporal difference learning computes, on each trial t, the difference δ_t between the outcome r_t and the expectation Q_t(action,state) for the choice at that state, and then uses this difference to update the expectation. Given these cached state-action Q values, choice is straightforward: simply choose the action that in the past yielded the best outcomes, i.e., the one with the largest Q value in that state. Since it is not

necessary to build or search a tree of future possibilities, these methods are computationally cheap. Indeed, further phases of model-free caching are also possible – towards storing the minimal information required, namely which action to take at each state (which then obviates the need to compare actions according to their values). In RL, such minimal models are called actors.

Model-free techniques replace the computational demands of model-based techniques with a demand for extensive experience. Critically therefore, because choices are based on long-term average returns, they are not sensitive to sudden changes in the values of outcomes, and so have been compared to habits (Daw et al., 2005). In Figure 1b, we see that after extensive experience, animals continue to press a lever for an outcome they no longer desire. In model-free accounts of choices, this is because choices are based on their past successes, not on whether they will now lead to the desired outcome. Thus, as experience accrues, organisms gradually shift control from goal-directed towards habitual choice, which is no longer directly driven by the immediate outcome (Dickinson and Balleine, 2002; Tricomi et al., 2009). Normative accounts argue that this shift occurs as the inaccuracies associated with the taxing computations of model-based control outweigh the inaccuracies associated with the statistical inefficiency of model-free control. It has been found, however, that both controllers are functional (Coutureau and Killcross, 2003; Smith et al., 2012) simultaneously, and various more complex interactions are being examined (Daw et al., 2011; Huys et al. 2012; Dayan 2012).

One important interaction is strongly predicted on theoretical grounds, but has yet to be observed in the wild. This is that cached values should be used to substitute for the values of whole branches of the goal-directed search tree, when that tree gets too big to evaluate directly. This is common in technical applications, underlying, for instance, the operation of Deep Blue, the first computer to beat a reigning chess world champion (Campbell et al., 2002). We will return to this in the section on conditioned reinforcement. Amongst other facets, it provides a means for comparatively inflexible habits to distort the flexible goal-directed system, and hence for the staggering contrast between the fixed desire for drugs, but the broadly adaptive strategies addicts characteristically engage in to obtain those drugs.

Instead of updating the expectation of state-action combinations, state expectations V(state) alone can also be maintained. These state expectations, which could also be produced by a model-based calculation, have been treated in Pavlovian terms, representing predictions about future prospects in an action independent manner. Such Pavlovian conditioning is evident in conditioned responses, namely actions elicited in states that predict affectively important outcomes, but whose form seems to be predetermined. States with positive values, i.e. that lead to reward, may elicit approach and engagement. States that lead to punishments evoke inhibition and withdrawal (McClure et al. 2003). The detailed nature of the responses evoked by valued states are thought to be evolutionarily pre-programmed and can be surprisingly fine-grained, particularly for the case of aversion (Bolles, 1970). However, because they are evolutionarily engrained, only certain state-action mappings can be implemented by the Pavlovian system (in contrast with the instrumental system, where any state-action combination can theoretically be reinforced). This, combined with the cached nature of state values, explains one signature feature, which is that they are emitted when a certain stimulus elicits them irrespective of the actual consequence. For instance, pigeons peck a light predictive of food (hence having high state value V), even though food is omitted whenever they peck the light (Williams and Williams (1969); Hershberger (1986); Breland and Breland (1961); Dayan et al. (2006)).

Electrophysiological recordings from midbrain dopamine neurons have shown that their phasic responses are commensurate with a temporal version of just the sort of prediction error δ_t (obtained minus expected reward) necessary for incremental model-free learning in the face of rewards. Figure 2 shows the evolution of the value signal and the prediction errors δ_t for a simple Pavlovian task in which a light predicts a reward (see Montague et al., 1996; Schultz et al., 1997). These prediction errors allow the iterative, gradual and model-free learning of state values V in Pavlovian conditioning (Montague et al., 1996; Schultz et al., 1997; Bayer and Glimcher, 2005), and of state-action values Q in instrumental conditioning (Schultz et al., 1997; Morris et al., 2006; Roesch et al., 2007) by simply accumulating them over time. It has been observed that the connections between the striatum and dopamine nuclei form a

spiral, running from ventromedial to dorsolateral striatum in consort with VTA to SNc (Haber et al., 2000; Joel and Weiner, 2000). This spiral has been suggested as being consistent with the progressive realization of the minimal form of model-free (habitual) control (i.e., an actor), in the most dorsal regions of the striatum (Haruno and Kawato, 2006).

Summary box 1: Phasic dopamine signals are thought to support the acquisition of model-free reward expectations and to drive habitual decision making. Model-based expectations underlie goal-directed choices. In Pavlovian paradigms, state or stimulus-evoked expectations themselves have a direct influence on behavior.

Clinical vignette: A patient presented to Accident & Emergency with signs of alcohol withdrawal. Inquiries with other local hospitals and the community mental health team revealed that he had presented to A&E with similar symptoms three times. Every of these times, he had undergone detoxification as an inpatient, in that the potentially severe side effects of sudden cessation of alcohol drinking had been managed with a tapered course of benzodiazepines. Moreover, each time he received detailed counselling about alcohol cessation and had expressed strong wishes to remain alcohol free. This was in part to save his job as a staff nurse and his marriage. On this, his fourth, admission, however, he lost his job and his wife had left him. On questioning, he explained how his relapses had always taken place after returning to his usual drinking places. He tried to avoid them. However, when he was tired or otherwise preoccupied, he would find himself habitually taking the turns leading him back to his favourite pub. Although manageable from afar, once close, he found the temptation to drink much more difficult to resist. Once he had resumed drink, he then found himself unable to stop. Each drink lead to the next, as if cueuing it, eventually culminating in the next admission to the hospital.

Towards "compulsions" in addiction

The power of some drugs of addiction to release dopamine directly (e.g. Aragona et al. 2008) allows them to act in place of natural appetitive reinforcers, directly boosting actions leading to the drugs (Phillips et al. 2003; Redish, 2004; Redish et al., 2008; though see also Self and Nestler, 1995). That is, without altering the function or structure of reward signaling or learning per se, drugs can usurp the learning structures by virtue of their impact on dopamine. In itself, this may lead to very strong behaviors and likely plays an important role both in the initial phases of drug taking and in the maintenance. Nevertheless, in a subset of the population, drug taking takes on a nature that appears to be more "compulsive" than merely habitual in that it becomes immune to aversive consequences. This is mirrored by findings in the animal literature where extensive experience with lever pressing for an addictive substance leads to responding that persists when shocks are superimposed (Vanderschuren and Everitt, 2004), but only in a subpopulation of subjects (Deroche-Gamonet et al., 2004). The mechanisms that differentiate habitual drug taking and facilitate their malignant transformation into more compulsive forms are the object of current research. We here review three important factors: first, adaptations in the dopaminergic system; second, adaptations in the striatal embedding of state-action contingencies; and third, Pavlovian state valuation processes (see also Huys et al., 2013).

Neuroadaptations of D_2 receptors

Drugs of abuse, particularly amphetamines, persistently modify the dopaminergic system, potentially increasing the amount of dopamine released over and above the physiological levels (Beck et al. 2011, Di Chiara and Bassareo, 2007; Heinz et al., 2009a). This is supported by a prominent collection of findings concerning the role of D_2 receptors in addiction. PET imaging in humans addicted to a variety of substances has shown reductions in striatal D_2 receptors (Heinz et al., 1996, 2009b; Volkow et al., 2009). There are parallel findings in animal models even in obesity (Johnson and Kenny, 2010). Reductions in D_2 receptor densities in humans may even predispose to the development of addictive behavior (Volkow et al., 2009; Morgan et al., 2002), and a causal role of such D_2 reductions for excessive drug intake been demonstrated causally in animal experiments (Thanos et al., 2001). In humans, the reduction correlates with subjective reports of craving (Heinz et al., 2004; Figure 3), likely increases the risk of relapse (Heinz et al., 2005b), and modulates the euphoria experienced by amphetamine administration (Volkow et al., 2002b). In the human midbrain, it correlated with increases in impulsivity (Buckholtz et al., 2010). Indeed, allelic variation in genes thought to affect the D_2 receptor point in the same direction (Comings et al., 1996; Frank et al., 2007).

One important limitation of PET studies is that they have had difficulty dissociating pre- and post-synaptic contributions of D_2 receptors. D_2 receptors have a high affinity for dopamine (Richfield et al., 1989). They are therefore substantially occupied even at tonic background levels of dopamine (up to 70%; Abi-Dargham et al., 2000; Laruelle et al., 1997; Dreyer et al., 2010). Since they are inhibitory (Nicola et al., 2000), those D_2 receptors located presynaptically might thus have a role in modulating phasic transients, i.e. in a negative feedback loop reducing phasic dopamine transients in response to tonic dopamine levels (Grace, 1991; Floresco et al., 2003; Goto et al. 2005; Niv et al., 2007; Schlagenhauf et al., 2012;). Changes in presynaptic D_2 receptors could contribute to an increase in phasic dopamine signals in at least two ways. First, a reduction in D_2 receptors would reduce the negative feedback, and would also weaken the connection between tonic and phasic signals (Lorenz, Schlagenhauf, Huys, Heinz, unpublished observations). Indeed, mice genetically modified to lack presynaptic D_2 receptors (but not differing in tonic dopamine levels) show an increase in phasic dopamine release (Bello et| al. 2011). Second, in the context of addiction, PET studies have also suggested a chronic reduction in released and stored dopamine (immediately post detoxification; Martinez et al., 2005; Heinz et al., 2005b; Volkow et| al. 1997, 2002a, 2007). If the indirect effect of reducing inhibition because the D_2 receptors will

experience reduced tonic dopamine levels outweighs the direct impact of the lessened release, this could also boost net phasic signals.

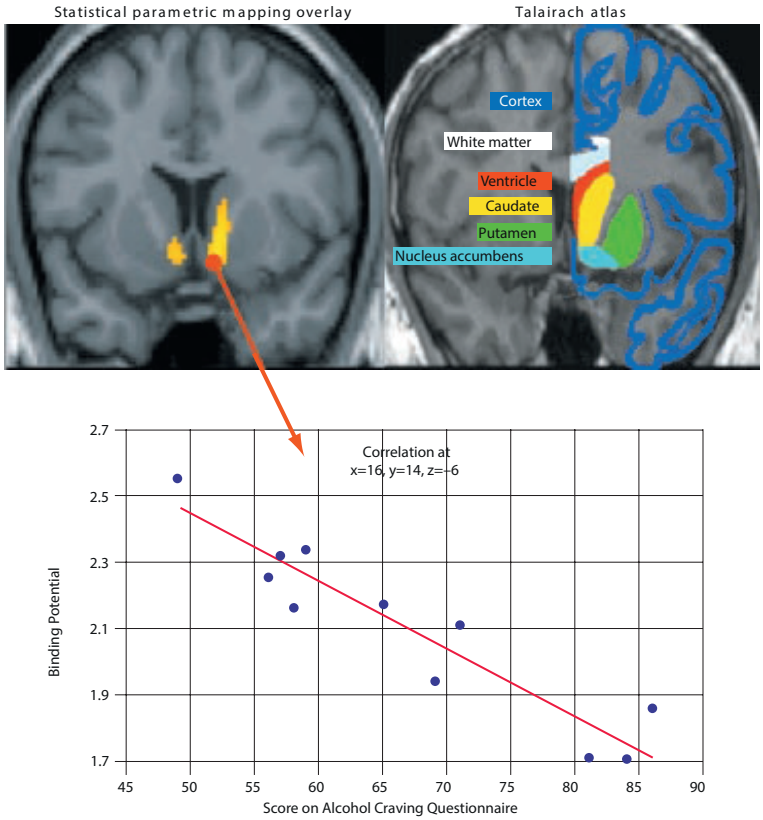

Figure 3: Craving and D_2 receptor availability in detoxified alcohol-dependent patients. The upper left image that D_2 receptor availability correlates with alcohol craving in the bilateral ventral striatum. The upper right image superimposes standardized anatomic brain regions. The scatterplot at the bottom shows the correlation between D_2 binding potential in the right ventral striatum and the alcohol craving score. Adapted from Heinz et al. 2004, permission pending. See original publication for further details.

Alterations of the generation and regulation of the phasic dopamine signal itself provide a direct pathway by which drugs of addiction might persistently modify the kinds of model-free learning signals (the prediction error δ_t) that dopamine is thought to report. The question is precisely how, as subtly different modifications make profoundly different predictions (Redish 2004; Dayan 2009), and some obvious modifications do not result in overly resistant behaviour. For instance, if the changes simply resulted in a multiplication of the prediction error signal, this would effectively lead to a change in the learning rate and predict faster learning, but the resulting behavior would have no more weight than other behaviors. A simple multiplicative increase in the phasic signal would not only fail to explain the persistence of drug taking against aversive consequences (as it would predict straightforward un-learning), but it additionally would fail to predict its dominance over other appetitively motivated behaviors. “Compulsive” behaviors could, however, emerge if the phasic dopamine signal was not multiplied, but rather boosted by a constant term. This would result in incessantly continuous reinforcement (Redish 2004). Although it is unclear whether a key prediction of this particular account holds (Iordanova et al., 2006; Panlilio et al., 2007; Jaffe et al., 2010), model-free RL provides a number of alternatives (Dayan, 2009), which have seen important elaborations and modifications (Dezfouli et al., 2009; Piray et al., 2010).

A further alternative is that the persistent alteration in postsynaptic D_2 receptors could specifically impact the learning of NoGo responses (i.e. learning not to emit a response) when less reward than expected is provided. This comes from the insight, worked out in impressive detail by Frank and colleagues (Frank and O'Reilly, 2006; Cohen and Frank, 2009; Frank, 2005; Frank et al., 2007; Frank and Claus, 2006), that D_2 receptors on neurons in the indirect pathway of the striatum play a prominent role in detecting pauses in the phasic activity of dopamine neuron (i.e., dips below baseline firing). Pauses come from negative prediction errors (i.e. when the obtained reward is smaller than expected). Because D_2 receptors have an inhibitory effect, pauses result in removing inhibition from the indirect pathway in a selective manner. In turn, this reduced inhibition allows the indirect pathway functionally to inhibit the actions that led to these dips and to facilitate learning not to execute them in the future (Shen et al., 2008). Indeed, recent optogenetic interrogation of these circuits has supported a role for the indirect, D_2-modulated system, in punishment and the associated inhibition (Kravitz et al. 2012). If the D_2 system is compromised, then subjects might be significantly less able to learn to suppress responses after punishments and avoid doing actions that ultimately turn out to be highly deleterious, such as those associated with drug taking (Vanderschuren et al., 2005; though see Piasecki et al., 2010).

Indeed, it is D_2 rather than D_1 agonists that enhance cocaine seeking and promote relapse (Self et al. 1996). Moreover, addictive and/or compulsive behaviours are well-described side-effects of D_2 agonist treatments in Parkinson's disease (Dagher and Robbins 2009). The involvement of both D_2 downregulation and overstimulation raises questions about the exact mechanism that links D_2 dysregulation with addiction. One explanation might be that D_2 downregulation is often observed in patients after detoxification, and hence may represent an adaptive consequence of pathologically high dopamine levels. A second possibility is a dissociation between pre- and postsynaptic D_2 effects. Indeed, because presynaptic effects mainly impact phasic dopamine signals, which in turn are sensed by D_1 receptors (Dreyer et al., 2010), this may also map on differences between direct and indirect pathways. Third, while continuous stimulation of D_2 receptors could prevent the learning contingent on dopamine dips, a downregulation of D_2 receptors could reduce the specificity with which aversive reinforcement is associated with events. This latter process may also involve dysfunctions in orbitofrontal cortex function (Takahashi et al., 2011), which are associated with D2 receptor changes in addiction (Volkow et al., 2007). Finally, synaptic plasticity at D_1 receptors may also contribute to addictive behavior (Pascoli et al. 2012).

In the following, we focus on two mechanisms that may be of particular importance: predominantly instrumental ones to do with neuroadaptations and deep striatal embedding, and a predominantly Pavlovian one to do with an enhanced effect of state values on behavior. These effects may be synergistic, and may even turn out to depend on related underlying neural substrates.

> Summary box 2: D_2 receptors in the striatum play a key role in the development of addiction as well as established addiction. One possibility is that they do this by influencing phasic dopamine signals. Alternatively, they might specifically impair Nogo-learning after punishments or non-rewards and thereby facilitate the persistence of addiction in the face of negative consequences.

Habitization and striatal embedding

An additional factor that might affect the inability to change behavior when shocks are superimposed is the progression mentioned above from goal-directed to habitual behavior, and, thereafter, to a minimal, value-independent, actor. As described above, model-based goal-directed behaviors dominate performance early in learning, while cached behaviors take over after extensive training.

The realization of the switch from goal-directed to habitual responding remains unclear, but in healthy individuals, prelimbic and infralimbic cortical regions appear to play a significant part (Killcross and Coutureau, 2003; Smith et al., 2012). Similar to experience, stress and alcohol are known to promote habitual over goal-directed responding and lead to the emergence of the signature outcome-insensitive

behavior (c.f. Figure 1b; Schwabe and Wolf, 2009, Ostlund and Balleine 2008). However, whether this involves the same mechanisms as in healthy states is unclear. A priori, it could reflect either a strong expression of habitual behavior, be driven by a weakening of the goal-directed controller, or represent a combination of both. While arguments that stress the influence of drugs of abuse on phasic dopamine signals favor the former (e.g. Redish, 2004), other accounts emphasize the latter. Ostlund and Balleine (2008), for instance, suggest that the shift from goal-directed to habitual responding observed with alcohol really should be seen as a failure by the goal-directed system to reassert its influence rapidly when change is evident. Indeed, there is direct experimental evidence that prelimbic cortex activity can ameliorate persistent drug seeking (Chen et al. 2013). The third option – a combination of both processes – is suggested by the finding that changes in striatal D_2 receptor in addiction correlate with changes in prefrontal function (Volkow et al., 2001, 2002a; Heinz et al., 2004; Park et al., 2010).

As discussed above, cached values are insensitive to a sudden change in outcomes, and can only reflect the summed past experience by slowly averaging over iterative updates. This is in contrast with goal-directed evaluations, which reflect novel information much more rapidly (Dickinson and Balleine, 2002; Daw et al., 2005; de Wit et al., 2009; Daw et al. 2011). Thus, if drug-taking actions are cached as highly appetitive, then they could be relatively less sensitive to the new-found delivery of shocks. However, even cached values are not eternally insensitive to the continued presence of aversive outcomes, but are amenable to further modification through experience: a drug habit is more robust to experience than the standard habits that ease our daily life. Thus, habitual responding as captured by cached values needs to be further fortified to turn into a "compulsion". During the natural course of prolonged learning, habits have been found to undergo just such a fortification, whereby the neural instantiation of habits migrates in loops. Early on, the correlates of behavior are prominent in ventral striatal areas, while later on after extensive training, the correlates of less flexible behavior are detectable in the dorsal striatum (Joel and Weiner, 2000; Haber et al., 2000; Yin et al., 2004; Belin et al., 2009). It is once again not clear that this process can lead to "compulsivity" with natural reinforcers, although Dayan (2009) describes two reinforcement learning algorithms, relying on actor-critic and advantage learning, that may allow for this.

Various findings suggest that the ventral to dorsal migration is an important process in the development of habits, and potentially of deeply engrained addictive behaviors. The shift depends on dopamine (Faure et al., 2005). It is reflected by a change in the pattern of dopamine release (ventral, then dorsal; Ito et al. 2002). Alcohol hastens the establishment of habits in rodents (Dickinson et al., 2002). In primates, functional changes progressing from ventral to dorsal striatum have also been observed in response to cocaine self-administeration (Porrino et al., 2004). Further, both dopaminergic lesions in the dorsolateral striatum (Vanderschuren et al., 2005) and disconnection of ventral to dorsal connectivity abolishes what may be a key link in the chain of the development of a "compulsion", namely drug seeking behavior (Belin and Everitt, 2008). However, the relationship of the dorsal migration to D_2 receptor abnormalities is not yet fully understood.

Summary box 3: Progressive embedding of habits from ventral to dorsal striatum via spiraling loops may provide a substrate for the "compulsive" nature of addictive behaviors.

Sign tracking

Ever since the seminal description of opiate addicted soldiers returning from Vietnam (Robins et al., 1974), environmental cues associated with drug taking have been thought likely to make a powerful and potentially perpetual contribution. In the laboratory, these effects can be examined in Pavlovian conditioning paradigms, where stimuli (e.g. a light) are presented at unpredictable times, but predict the imminent delivery of a reward (say a food pellet or an addictive drug). Responses to drug-associated conditioned cues have long been examined in humans and shown strong effect sizes. Patients with a variety of drug addictions report more craving and show more physiological responses when seeing drug-associated cues (Carter and Tiffany, 1999). Furthermore, there is evidence that the prefrontal response to passively viewed drug-associated cues is affected by striatal D_2 receptors (Heinz et al., 2004), and that these responses are predictive of relapse (Grüsser et al., 2004).

Theoretical accounts of Pavlovian learning, i.e. of the establishment of such responses, emphasize the attribution of a value to a stimulus or situation. This is in contrast to instrumental learning, which attributes value to state-action pairs. However, as we mentioned, similar to the instrumental state-action case, values of states can be derived by model-based or model-free mechanisms. While the latter are thought to relate to iterative, trial-and-error learning via a dopaminergic prediction error, the status of the former are less certain, but, if they are shared with instrumental model-based evaluation (as they may do only partially; Robinson & Berridge, 2013), they likely involve the prefrontal cortex – the prelimbic cortex in rodents and the ventrolateral prefrontal cortex as the human analogue (Schoenbaum et al., 2009; Walton et al., 2010; Takahashi et al., 2011).

As we discussed, in Pavlovian conditioning, subjects may emit a number of different behaviors in response to the presence or predictions of the presence of reinforcers (i.e. the value) (Devenport 1979). Unlike the instrumental case, the form of the emitted behavior is linked inflexibly (likely via evolutionary mechanisms) to the stimulus. The value attributed to the stimulus determines whether the behavior is emitted and how strongly. Because the identity of the behavior is linked to the stimulus rather than to the achievement of a goal or reception of a reward, these Pavlovian responses are likely adaptive on an evolutionary scale, but have the potential of being counter-productive in specific instances (Boakes, 1977; Hirsch and Bolles, 1980; Hershberger, 1986; Williams and Williams, 1969; Dayan et al., 2006, Guitart-Masip et al. 2012).

Flagel and colleagues have recently examined individual variation in the situation where a light bulb predicts a food pellet, but where the light bulb is mounted at a different spatial location to where the food is being delivered (Meyer et al., 2012). Some rats, henceforth called 'goal trackers', come to immediately move towards the food pellet delivery site when the light comes on. Others, the 'sign trackers' instead approach the light (the sign; Figure 4A,D). In both groups the light, which has come to predict an appetitive event, modified behavior. Only the sign trackers, however, emit the response of approaching the predictor. This property of the sign is called incentive salience by Berridge and Robinson (1998).

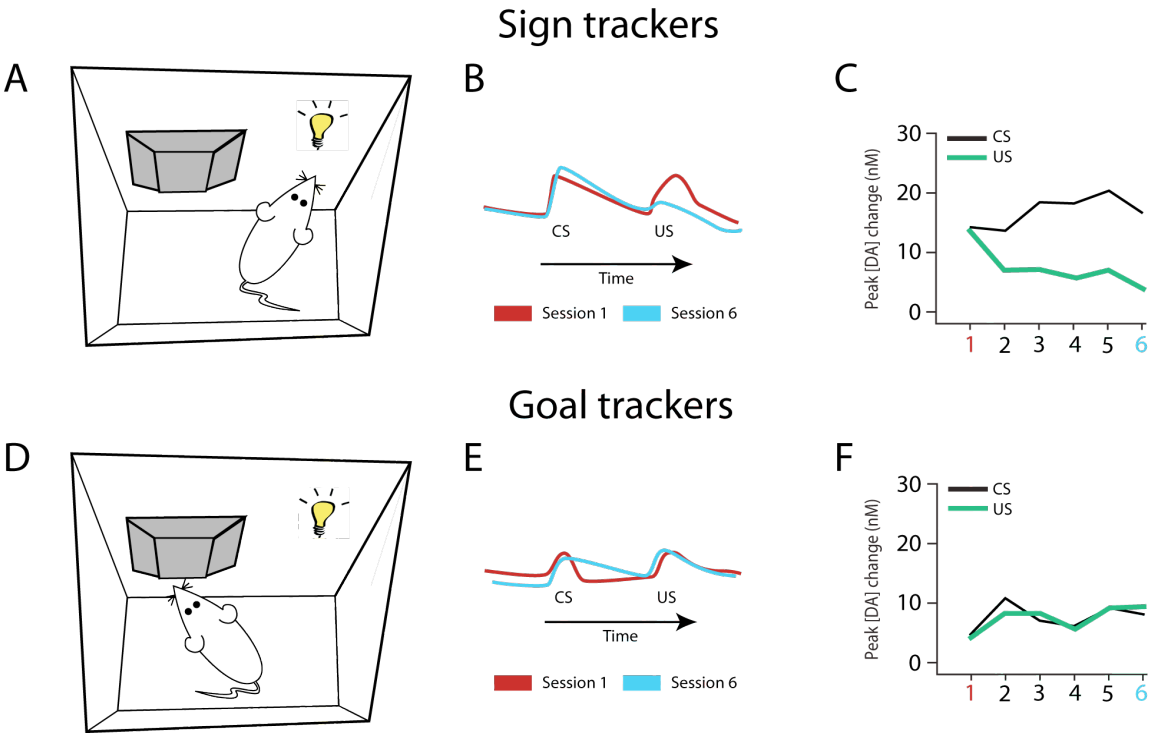

Figure 4: Sign tracking and goal tracking by rats exposed to classical conditioning, whereby a light signal predicts delivery of a food US at a different location. A&D: Sign tracking rats come to approach the light during the delay between US and CS, while goal-tracking rats

approach the location where the food US will be delivered. B&E: Phasic dopamine signals in the nucleus accumbens core. In sign trackers, the phasic response to the CS increases, while that to the US decreases, just as predicted by the temporal prediction error hypothesis (Figure 2B). In Goal trackers, phasic dopamine responses to CS and US do not change over time. C&F show how the peak dopamine responses change over trials. These differences suggest that sign trackers acquire a cached value V in accordance with the temporal prediction hypothesis, but that goal trackers do not. Data in B,C,E,F adapted from Flagel et al. 2011.

In sign trackers, dopamine released in the nucleus accumbens core (a key region for Pavlovian effects on behavior; Holmes et al. 2010) shows the the characteristic patterns of a temporal difference prediction error (Figure 4B,C; see Huys et al., subm. for a detailed discussion): over time, the phasic response to the onset of the CS increases, while the phasic dopamine response to the onset of the US decreases. In goal trackers, on the other hand, the phasic dopamine signals do not adapt, as if the cached signal was not being used to acquire the value of the signal (Figure 4E,F). Thus, the phasic dopaminergic signals only resemble the model-free prediction error signals in one group: the sign trackers. Maybe even more strikingly, it is also only in the sign trackers that a dopaminergic antagonist abolishes the response to the conditioned cue (Flagel et al., 2011). That is, these phasic dopaminergic signals appear only to be critical for learning in the sign trackers. Furthermore, animals that sign track to food (unlike goal trackers) have a distinctively 'addictive' neuropsychological profile: they also sign track to cocaine (Flagel et al., 2009), are more susceptible to the sensitizing effects of amphetamine (Flagel et al., 2008), and they are more novelty seeking (Flagel et al., 2010) and impulsive (Lovic et al., 2011). The boosted dopamine signal is reminiscent of the finding discussed earlier whereby reductions in pre-synaptic D_2 receptor sensitivities boost phasic dopamine signals (Bello et al., 2011). In fact, animals that sign track do show lower D_2 receptor densities in midbrain dopaminergic areas (Flagel et al., 2007). Finally, while sign-tracking and goal-tracking sub-populations exist in outbred strains, there is also a genetic influence as these phenotypes can be bred true (Flagel et al., 2011).

Thus, it appears that only one group of subjects, the sign trackers, shows cached Pavlovian conditioning maintained by iterative updates via phasic dopaminergic prediction errors; and that those subjects prone to employing a cached learning mechanism are more vulnerable to develop addictive behaviors. The conditioning observed in the goal trackers could well be a signature of goal-directed behavior, being independent of phasic dopaminergic prediction errors (Dickinson et al., 2000; Wassum et al., 2011). In the words of the authors, goal-trackers do not assign 'incentive value' to the predictive stimuli (McClure et al., 2003; Huys et al., subm.); a derivation of value via a goal-directed mechanism would predict selective sensitivity to outcome devaluation in goal trackers (Allman et al. 2010). We note that the addiction vulnerability of subjects who use a cached Pavlovian value to guide action choice might possibly be another facet of the shift from model-based, or goal-directed, to model-free, or cached, decision making discussed above. Indeed, the fact that sign-trackers do not develop a goal-tracking response under DA antagonism suggests that in the absence of iterative caching mechanisms, these animals have no alternative backup learning mechanism or are unable to engage it. Given reports of parallel representations of habitual and goal-directed responding (Coutureau and Killcross, 2003), this points to a parallel impairment in goal-directed function, possibly due to variations in prefrontal function (Volkow et al., 2009).

Robinson and Berridge (2003, 2001) suggest the additional possibility of incentive sensitization – namely that the dopamine release associated with drug-associated cues is boosted to abnormal levels by a direct physiological adaptation to the drug delivery. This could emphasize those cues, and lead to untoward actions under Pavlovian control, even if maladaptive. However, this has to date not been shown in humans (Heinz et al., 2004; 2005b; Volkow et al., 1996).

It is worth noting an additional possible effect of internal states on relapses. Many drugs exert their strongest immediate impact after a period of abstinence (rather as food is often more tasty when you are hungry). If the state of abstinence can be internally recognized, then the value of cues associated with

drugs (i.e., the predicted value of the drugs to which they lead) when 'clean' could be particularly high, and thus particularly hard to resist. In fact, this problem can afflict instrumental mechanisms too.

Summary box 4: Animals whose phasic dopaminergic signals in the ventral striatum behave as if they were involved in acquisition of cached values show addictive traits.

Conditioned reinforcement

A final effect that we consider is the possibility suggested by the sign trackers, that vulnerabilities to the development of addiction may lie in the extent to which cached values influence behavior. The nucleus accumbens core is critical to another aspect of Pavlovian influence on behavior (Parkinson et al., 1999), namely second-order conditioning. In this, subjects work not to obtain access to drugs or other direct reinforcers themselves (USs), but rather to obtain access to conditioned stimuli (CSs) that have previously been associated with the USs. In other words, it describes the process by which neutral CSs can come to motivate behavior in a manner akin to USs. Animal experiments showed that CSs predictive of drugs are extremely stable and powerful in supporting second-order conditioning, even long after the association between the CS and the US has been extinguished (Di Ciano and Everitt, 2004). However, it is again only sign trackers that show this behavior (Robinson and Flagel, 2009). If we follow Everitt and Robbins (2005) in interpreting drug seeking as conditioned reinforcement, then this suggests that drug seeking would preferentially occur amongst subjects whose dopaminergic system builds cached values.

The mechanisms of second-order conditioning may play an additional role. As we mentioned above, when goal-directed search trees become too large to build and search, it is expected that cached values will be used to substitute for the evaluation of extensive branches of the tree. For instance, the queen is the most powerful figure in chess. Winning without a queen is hard, and manoeuvres that 'sacrifice' the queen in order to achieve a win are rare and viewed as very elegant. This may be because rather than thinking through the consequences of sacrificing the queen, players are dissuaded from sacrificing the queen by its cached value, acquired over previous experience (Huys et al., 2012). This substitution of subcomponents of the goal-directed decision process by cached values is one possible avenue to second-order conditioning, but acting between the model-based and model-free systems. The resulting hybrid could provide an account for why drugs addicts are on the one hand able to engage in highly complex and goal-directed drug seeking, but then entirely fail to consider the outcomes of drug taking: they use complex, goal-directed tree search to seek the drug state; but the value of the drug state itself is model-free. It is therefore not further decomposed and its subcomponents not accessible to the goal-directed system. High-level cognition is therefore distorted by the habits but cannot modify or comprehend the components making up the value of the habit itself.. By the same token, interrupting the Pavlovian attraction to drug related stimuli (e.g. by attention control training; Fadardi and Cox 2009; Wiers et al. 2011) may interrupt the maintenance or further development of the habit / "compulsion".

Conclusion

The power of addictive drugs to seize control over the apparatus of decision making in vulnerable individuals is awesome and awful in equal measure. It turns out that the key problem to resolve is not so much the initiation, but rather the development of the disorder, leading as it does, in a sub-population, to "compulsive" behavior that resists being curtailed by the obviously evident negative outcomes.

The field of neural reinforcement learning is providing a wide range of both quantitative and qualitative foundations for encompassing the wealth of data, and providing foundations for some of the implementational, algorithmic and computational flaws that might be responsible (Redish, 2004; Redish et al., 2008; Dezfouli et al., 2009; Piray et al., 2010; Dayan, 2009; Huys et al., in press). We have discussed a number of instrumental and Pavlovian possibilities; these are covered in more detail in the references. We paid particular attention to model-based and model-free forms of instrumental behaviour, along with Pavlovian influences. Amongst various distinctions, it is likely that model-free forms of prediction and control are largely hidden from explicit subjective view. However, the nature of subjective access to model-based prediction and control, and what happens when model-based and model-free systems interact (as when values from the latter are used to substitute for branches in the tree-based evaluation of the former) are as yet incompletely clear. Conversely, among things that these systems are all expected to exploit is the representation of the decision-making domain. Such a representation forms the foundation for all predictions, and its own realization poses substantial computational (and statistical) demands. Since representations are in common between the systems, we did not discuss them here; however, it is recently being recognized that many phenomena in normal and additive decision-making such as extinction and spontaneous recovery therefrom may be understood through an analysis of how different cases of learning lead to generalizing or specializing representations (Courville et al., 2004; Redish et al., 2008; Gershman et al, 2010).

We focused on dopamine, because it appears critical for the common mode of action of drugs of addiction. However, other neuromodulators and neurotransmitters also appear to play critical roles in some aspects of the development and maintenance of addictions. For instance, opioids have been discussed by Berridge (2009) as being involved in the process of hedonic evaluation of outcomes (and thus cues leading to those outcomes). Such hedonic effects may influence learning directly, and could also modulate RL mechanism by changing apparent outcome values. This offers an alternative route to the sort of incentive salience or even incentive sensitization discussed above (Berridge and Robinson (1998); Robinson and Berridge (2003)). For instance, one PET study revealed that μ-opioid receptors are up-regulated in detoxified alcohol-dependent patients compared to healthy controls, and that the degree of up-regulation directly correlated with alcohol craving. Alcohol intake in such patients may then release endorphins and induce increased pleasure. Indeed, opiate antagonists can be used clinically, and some patients in the study reported that the desire to consume alcohol during a relapse was reduced when they took naltrexone, a drug that blocks opiate receptors including the mu-opiate receptors in the ventral striatum (Heinz et al., 2005a, 2009a).

In addition, there is ample speculation about the role of serotonin as an opponent to dopamine, with mechanisms by which both increases and decreases in serotonin concentration can impact the evaluation of states and outcomes (Cools et al., 2009; Daw et al., 2002; Dayan and Huys, 2008, 2009; Boureau and Dayan, 2011; Cools et al., 2011). There is evidence that serotonin transporters are impaired during alcohol detoxification (Heinz et al., 1998), and this could, for instance, interfere with limbic activation to threatening stimuli and contribute to negative mood.

Perhaps the most important immediate direction for the next phase of modeling is to incorporate the significant insights available from the extensive data on the effects of neurobiological markers such as D_2 receptor function and vulnerabilities to turning casual drug use into a fully-fledged addiction. This is

nicely consonant with recent trends in the research of decision making to embrace the richness and complexity of individual differences.

One aspect of this research that requires further exploration is the notion of "compulsions". The term compulsion was developed within the theoretical framework of OCD to denote stereotypic acts that are meant to counteract obsessive thoughts – e.g. aggressive, blasphemic or involving contamination. Since thoughts can recur repeatedly, so do compulsions. Compulsions in OCD appear to be senseless to the individual who nevertheless can hardly refrain from enacting them. Interruption of the rituals results in anxieties because they are believed to be necessary to counteract or avoid the obsessions. Compulsions are repeated incessantly, devoid of flexibility, accompanied by negative mood states if interrupted and accompanied by an inner sense of lack of closure (Heinz, 1999). This contrasts with habitual acts of drug taking which are hardly ever so stereotypic or pervasive. Interruption of habitual drug taking results in craving (Carter and Tiffany, 1999) rather than anxieties, and it is the craving that dominates the subjective sensation when engaging in drug seeking. Furthermore, unlike with compulsions in OCD, "compulsions" in the context of addiction are not completely irresistible (drug habits can be 'kicked' without specific therapy); rather, they reflect a fundamental biasing of behavior towards drug use. Nevertheless both syndromes show a certain behavioural inflexibility and insensitivity to outcomes, and RL might have a role in describing how these two features arise. We have discussed how RL can help to explain how motivational mechanisms are profoundly affected by chronic drug use and propel an individual towards inflexible or maladaptive behaviour despite aversive consequences. They might also capture learning processes in the context of OCD, such as the reinforcement of certain rituals by a reduction in anxiety (so-called avoidance learning; Maia 2010; Moutoussis et al., 2008). To what extent this kind of learning - the avoidance of an aversive state in OCD - is subserved by similar or different neural processes as the engraining of appetitively motivated behaviours in the initial phases of addictions is to be explored.

A further point worth considering is the relationship between addictive "compulsions" and passions. Both are characterized by engulfing desires that can drive a person to break through the barriers of social rules and norms (Plessner 2003). In terms of defining individual preferences, passions sit comfortably with the initiation phase of addictions, although passions are multifold and can persist without ever being fulfilled, which differs from common forms of drug craving and consumption. Also, a conflict between an environment and an individual's passions is by itself no useful criterion to distinguish a mental malady: social rules are so diverse as to make this a non-definition (consider the social norms relating to alcohol or consensual sexual practices). Thus, the fact that drug consumption has certain adverse social consequences is not by itself sufficient to either warrant the label "compulsion" or to define a psychiatric disorder. Rather, it is the pervasive loss of flexible behaviour, adaptation and control that represents an impairment of a relevant higher cognitive function and can thus serve as a criterion for diagnosing a mental illness (Heinz & Kluge 2010).

In sum, together with the models for the initiation of addiction that are well described (Redish, 2004; Redish et al., 2008; Dezfouli et al., 2009; Dayan, 2009; Gutkin and Ahmed, 2011), neural RL provides many routes to the formalization of the phenomenological experience of initiation, craving, loss of control over intake of drugs, habitization of drug taking and even the manifestation of anxiety and arousal during withdrawal. Computational accounts of drug addiction started from a simple observation about the devastating consequences of ectopically seizing the reigns of the prediction error; however through theory and models, this domain of enquiry has evolved to being richly revealing about the overall architecture of choice.

Funding & conflicts of interest

QJMH, AH and AB were supported by the German Research Foundation (DFG FOR RA-1047/2-1). PD was supported by the Gatsby Foundation. The authors report no conflicts of interest.

References

Abi-Dargham, A., Rodenhiser, J., Printz, D., Zea-Ponce, Y., Gil, R., Kegeles, L. S., Weiss, R., Cooper, T.B., Mann, J.J., Van Heertum, R.L., Gorman, J.L. and Laruelle, M. (2000). Increased baseline occupancy of D2 receptors by dopamine in schizophrenia. Proc. Nat. Acad. Sci., 97(14), 8104-8109.

Allman, M. J., DeLeon, I. G., Cataldo, M. F., Holland, P. C., and Johnson, A. W. (2010). Learning processes affecting human decision making: An assessment of reinforcer-selective pavlovian-to-instrumental transfer following reinforcer devaluation. J Exp Psychol Anim Behav Process, 36(3):402–408.

American Psychiatric Association (1994). Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Association Press.

Aragona, B. J.; Cleaveland, N. A.; Stuber, G. D.; Day, J. J.; Carelli, R. M. & Wightman, R. M. (2008). Preferential enhancement of dopamine transmission within the nucleus accumbens shell by cocaine is attributable to a direct increase in phasic dopamine release events. J Neurosci, , 28, 8821-8831

Bayer, H. M. and Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron, 47(1):129–141.

Baxter, L.R. Jr., Schwartz, J.M., Mazziotta, J.C., Phelps, M.E., Pahl, J.J., Guze, B.H., Fairbanks, L. (1988). Cerebral glucose metabolic rates in nondepressed patients with obsessive-compulsive disorder. Am J Psychiatry. 145(12):1560-3.

Beck, A., Grace, A. A., Heinz A. (2011). Reward Processing. In: Adinoff, B.; Stein, E. (Eds.) Neuroimaging in the Addictions. John Wiley & Sons, New York, pp. 107-129.

Belin, D. and Everitt, B. J. (2008). Cocaine seeking habits depend upon dopamine-dependent serial connectivity linking the ventral with the dorsal striatum. Neuron, 57(3):432–441.

Belin, D., Jonkman, S., Dickinson, A., Robbins, T. W., and Everitt, B. J. (2009). Parallel and interactive learning processes within the basal ganglia: relevance for the understanding of addiction. Behav Brain Res, 199(1):89–102.

Bello, E. P., Mateo, Y., Gelman, D. M., Noaín, D., Shin, J. H., Low, M. J., Alvarez, V. A., Lovinger, D. M., and Rubinstein, M. (2011). Cocaine supersensitivity and enhanced motivation for reward in mice lacking dopamine d(2) autoreceptors. Nat Neurosci, 14(8):1033–1038.

Berridge, K. C. (2009). 'liking' and 'wanting' food rewards: brain substrates and roles in eating disorders. Physiol Behav, 97(5):537–550.

Berridge, K. C. and Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res. Rev., 28(3):209–69.

Boakes, R. (1977). Performance on learning to associate a stimulus with positive reinforcement. pages 67–101. Lawrence Erlbaum Associates, Inc. Hillsdale, NJ.

Bolles, R. C. (1970). Species-specific defense reactions and avoidance learning. Psychol Rev, 77:32–48.

Boureau, Y.-L. and Dayan, P. (2011). Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology, 36(1):74–97.

Breland, K. and Breland, M. (1961). The misbehavior of organisms. American Psychologist, 16(9):681–84.

Buckholtz, J. W., Treadway, M. T., Cowan, R. L., Woodward, N. D., Li, R., Ansari, M. S., Baldwin, R. M., Schwartzman, A. N., Shelby, E. S., Smith, C. E., Kessler, R. M., and Zald, D. H. (2010). Dopaminergic network differences in human impulsivity. Science, 329(5991):532.

Campbell, M., Hoane, A., et al. (2002). Deep blue. Artificial Intelligence, 134(1-2):57–83.

Carter, B. L. and Tiffany, S. T. (1999). Meta-analysis of cue-reactivity in addiction research. Addiction, 94(3):327–340.

Chen, B.T., Yau, H., Hatch, C., Kusumoto-Yoshida, I., Cho, S. L., Hopf, W. and Bonci, A. (2013) Rescuing

cocaine-induced prefrontal cortex hypoactivity prevents compulsive cocaine seeking. Nature, 496: 395.

Cohen, M. X. and Frank, M. J. (2009). Neurocomputational models of basal ganglia function in learning, memory and choice. Behav Brain Res, 199(1):141–156.

Comings, D. E., Rosenthal, R. J., Lesieur, H. R., Rugle, L. J., Muhleman, D., Chiu, C. and Gade, R. (1996). A study of the dopamine D2 receptor gene in pathological gambling. Pharmacogenetics and Genomics, 6(3), 223-234.

Cools, R., Frank, M. J., Gibbs, S. E., Miyakawa, A., Jagust, W., and D'Esposito, M. (2009). Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. J Neurosci, 29(5):1538–1543.

Cools, R., Nakamura, K., and Daw, N. D. (2011). Serotonin and dopamine: unifying affective, activational, and decision functions. Neuropsychopharmacology, 36(1):98–113.

Courville, A. C.; Daw, N.; Gordon, G. J. & Touretzky, D. S. (2004). Model Uncertainty in Classical Conditioning In: Thrun, S.; Saul, L. & Schölkopf, B. (Eds.) Advances in Neural Information Processing Systems 16, MIT Press.

Coutureau, E. and Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. Behav. Brain Res., 146(1-2):167–74.

Dagher, A. And Robbins, T. W. (2009). Personality, addiction, dopamine: insights from Parkinson's disease. Neuron, 61, 502-510.

Daw, N. D. and Doya, K. (2006). The computational neurobiology of learning and reward. Curr Opin Neurobiol, 16(2):199–204.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. Neuron, 69(6):1204–1215.

Daw, N. D., Kakade, S., and Dayan, P. (2002). Opponent interactions between serotonin and dopamine. Neural Networks, 15:603–16.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci, 8(12):1704–1711.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. and Dolan, R. J. (2011) Model-based influences on humans' choices and striatal prediction errors. Neuron, 69, 1204-1215.

Dayan, P. (2009). Dopamine, reinforcement learning, and addiction. Pharmacopsychiatry, 42 Suppl 1:S56–S65.

Dayan, P. (2012). How to set the switches on this thing. Curr Opin Neurobiol, ,22, 1068-1074.

Dayan, P. and Huys, Q. J. M. (2008). Serotonin, inhibition, and negative mood. PLoS Comput Biol, 4(2):e4.

Dayan, P. and Huys, Q. J. M. (2009). Serotonin in affective control. Annu Rev Neurosci, 32:95–126.

Dayan, P., Niv, Y., Seymour, B., and Daw, N. D. (2006). The misbehavior of value and the discipline of the will. Neural Netw, 19(8):1153–1160.

de Wit, S., Corlett, P. R., Aitken, M. R., Dickinson, A., and Fletcher, P. C. (2009). Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. J Neurosci, 29(36):11330–11338.

Deroche-Gamonet, V., Belin, D., and Piazza, P. V. (2004). Evidence for addiction-like behavior in the rat. Science, 305(5686):1014–1017.

Devenport, L. D. (1979). Superstitious bar pressing in hippocampal and septal rats. Science, 205(4407):721–3.

Dezfouli, A., Piray, P., Keramati, M. M., Ekhtiari, H., Lucas, C., and Mokri, A. (2009). A neurocomputational model for cocaine addiction. Neural Comput, 21(10):2869–2893.

Di Ciano, P. and Everitt, B. J. (2004). Conditioned reinforcing properties of stimuli paired with self-administered cocaine, heroin or sucrose: implications for the persistence of addictive behaviour. Neuropharmacology, 47 Suppl 1:202–213.

Di Chiara, G., Bassareo, V. (2007). Reward system and addiction: what dopamine does and doesn't do. Curr Opin

Pharmacol. 7:69-76.

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 308(1135):67–78.

Dickinson, A. and Balleine, B. (2002). The role of learning in the operation of motivational systems. In Gallistel, R., editor, Stevens' handbook of experimental psychology, volume 3, pages 497–534. Wiley, New York.

Dickinson, A., Smith, J., and Mirenowicz, J. (2000). Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists. Behav Neurosci, 114(3):468–483.

Dickinson, A., Wood, N., and Smith, J. W. (2002). Alcohol seeking by rats: action or habit? Q J Exp Psychol B, 55(4):331–348.

Dreyer, J. K., Herrik, K. F., Berg, R. W., & Hounsgaard, J. D. (2010). Influence of phasic and tonic dopamine release on receptor activation. J. Neurosci., 30(42), 14273-14283.

Everitt, B. J., Belin, D., Economidou, D., Pelloux, Y., Dalley, J. W., and Robbins, T. W. (2008). Neural mechanisms underlying the vulnerability to develop compulsive drug-seeking habits and addiction. Philos Trans R Soc Lond B Biol Sci, 363(1507):3125–3135.

Everitt, B. J. and Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nat Neurosci, 8(11):1481–1489.

Fadardi, J. S. and Cox, W. M. (2009). Reversing the sequence: reducing alcohol consumption by overcoming alcohol attentional bias. Drug Alcohol Depend, 101(3):137–145.

Faure, A., Haberland, U., Condé, F., and Massioui, N. E. (2005). Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. J Neurosci, 25(11):2771–2780.

Flagel, S. B., Akil, H., and Robinson, T. E. (2009). Individual differences in the attribution of incentive salience to reward-related cues: Implications for addiction. Neuropharmacology, 56 Suppl 1:139–148.

Flagel, S. B., Clark, J. J., Robinson, T. E., Mayo, L., Czuj, A., Willuhn, I., Akers, C. A., Clinton, S. M., Phillips, P. E. M., and Akil, H. (2011). A selective role for dopamine in stimulus-reward learning. Nature, 469(7328):53–57.

Flagel, S. B., Robinson, T. E., Clark, J. J., Clinton, S. M., Watson, S. J., Seeman, P., Phillips, P. E. M., and Akil, H. (2010). An animal model of genetic vulnerability to behavioral disinhibition and responsiveness to reward-related cues: implications for addiction. Neuropsychopharmacology, 35(2):388–400.

Flagel, S. B., Watson, S. J., Akil, H., and Robinson, T. E. (2008). Individual differences in the attribution of incentive salience to a reward-related cue: influence on cocaine sensitization. Behav Brain Res, 186(1):48–56.

Flagel, S. B., Watson, S. J., Robinson, T. E., and Akil, H. (2007). Individual differences in the propensity to approach signals vs goals promote different adaptations in the dopamine system of rats. Psychopharmacology (Berl), 191(3):599–607.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J Cogn Neurosci, 17(1):51–72.

Frank, M. J. and Claus, E. D. (2006). Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. Psychol Rev, 113(2):300–326.

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., and Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci U S A, 104(41):16311–16316.

Frank, M. J. and O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. Behav Neurosci, 120(3):497–517.

Garris, P. A., Kilpatrick, M., Bunin, M. A., Michael, D., Walker, Q. D. and Wightman, R. M. (1999). Dissociation of dopamine release in the nucleus accumbens from intracranial self-stimulation. Nature, 398(6722), 67-69.

Gelder, M., Harrison, P., and Cowen, P. (2006). Shorter Oxford Textbook of Psychiatry. Oxford University Press, Oxford, UK.

Gershman, S. J.; Blei, D. M. & Niv, Y. (2010). Context, learning, and extinction. Psychol Rev, 2010, 117, 197-209

Grace, A. A. (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. Neuroscience, 41(1), 1-24.

Grüsser, S. M., Wrase, J., Klein, S., Hermann, D., Smolka, M. N., Ruf, M., Weber-Fahr, W., Flor, H., Mann, K., Braus, D. F., and Heinz, A. (2004). Cue-induced activation of the striatum and medial prefrontal cortex is associated with subsequent relapse in abstinent alcoholics. Psychopharmacology (Berl), 175(3):296–302.

Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, L., Dayan, P., Duzel, E. and Dolan, R. J. (2012) Go and no-go learning in reward and punishment: interactions between affect and effect. Neuroimage, 62, 154-166.

Gutkin, B. and Ahmed, S., editors (2011). Computational Neuroscience of Drug Addiction. Springer.

Haber, S. N., Fudge, J. L., and McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. J Neurosci, 20(6):2369–2382.

Haruno, M. and Kawato, M. (2006). Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fmri examination in stimulus-action-reward association learning. Neural Netw, 19(8):1242–1254.

Heinz, A. (1999). Neurobiological and anthropological aspects of compulsions and rituals. Pharmacopsychiatry. 32(6):223-9.

Heinz, A., Beck, A., Grüsser, S. M., Grace, A. A., and Wrase, J. (2009a). Identifying the neural circuitry of alcohol craving and relapse vulnerability. Addict Biol, 14(1):108–118.

Heinz, A., Beck, A., Wrase, J., Mohr, J., Obermayer, K., Gallinat, J., and Puls, I. (2009b). Neurotransmitter systems in alcohol dependence. Pharmacopsychiatry, 42 Suppl 1:S95–S101.

Heinz, A., Dufeu, P., Kuhn, S., Dettling, M., Gräf, K., Kürten, I., Rommelspacher, H., and Schmidt, L. G. (1996). Psychopathological and behavioral correlates of dopaminergic sensitivity in alcohol-dependent patients. Arch Gen Psychiatry, 53(12):1123–1128.

Heinz A, Kluge A:Anthropological and Evolutionary Concepts of Mental Disorders. Journal of Speculative Philosophy. 24(3):292-307 (2010

Heinz, A., Ragan, P., Jones, D. W., Hommer, D., Williams, W., Knable, M. B., Gorey, J. G., Doty, L., Geyer, C., Lee, K. S., Coppola, R., Weinberger, D. R., and Linnoila, M. (1998). Reduced central serotonin transporters in alcoholism. Am J Psychiatry, 155(11):1544–1549.

Heinz, A., Reimold, M., Wrase, J., Hermann, D., Croissant, B., Mundle, G., Dohmen, B.M., Braus, D.F., Schumann, G., Machulla, H.J., Bares, R., Mann, K. (2005a). Correlation of stable elevations in striatal mu-opioid receptor availability in detoxified alcoholic patients with alcohol craving: a positron emission tomography study using carbon 11-labeled carfentanil. Arch Gen Psychiatry. 62(1):57-64.

Heinz, A., Siessmeier, T., Wrase, J., Buchholz, H.G., Gründer, G., Kumakura, Y., Cumming, P., Schreckenberger, M., Smolka, M.N., Rösch, F., Mann, K., Bartenstein, P. (2005b). Correlation of alcohol craving with striatal dopamine synthesis capacity and D2/3 receptor availability: a combined [18F]DOPA and [18F]DMFP PET study in detoxified alcoholic patients. Am J Psychiatry. 162(8):1515-20.

Heinz, A., Siessmeier, T., Wrase, J., Hermann, D., Klein, S., Grüsser, S. M., Grüsser-Sinopoli, S. M., Flor, H., Braus, D. F., Buchholz, H. G., Gründer, G., Schreckenberger, M., Smolka, M. N., Rösch, F., Mann, K., and Bartenstein, P. (2004). Correlation between dopamine d(2) receptors in the ventral striatum and central processing of alcohol cues and craving. Am J Psychiatry, 161(10):1783–1789.

Hershberger, W. A. (1986). An approach through the looking-glass. Anim. Learn. Behav., 14:443–51.

Hirsch, S. and Bolles, R. (1980). On the ability of prey to recognize predators. Z. Tierpsychol, 54:71–84.

Holmes, N. M., Marchand, A. R., and Coutureau, E. (2010). Pavlovian to instrumental transfer: a neurobehavioural perspective. Neurosci Biobehav Rev, 34(8):1277–1295.

Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P. and Roiser, J. P. (2012). Bonsai trees in your head: How the Pavlovian system sculpts goal-directed choices by pruning decision trees. PLoS Comput Biol, ,8, e1002410.

Huys, Q. J. M., Tobler P. N., Hasler, G., Flagel S. B. (in press). The role of learning-related dopamine signals in addiction vulnerability. Prog. Neurobiol.

Iordanova, M. D., Westbrook, R. F., and Killcross, A. S. (2006). Dopamine activity in the nucleus accumbens modulates blocking in fear conditioning. *Eur J Neurosci,* 24(11):3265–3270.

Ito, R., Dalley, J. W., Robbins, T. W., and Everitt, B. J. (2002). Dopamine release in the dorsal striatum during cocaine-seeking behavior under the control of a drug-associated cue. J Neurosci, 22(14):6247–6253.

Jaffe, A., Gitisetan, S., Tarash, I., Pham, A., and Jentsch, J. (2010). Are nicotine-related cues susceptible to the blocking effect? Soc. Neurosci. Abstr.

Joel, D. and Weiner, I. (2000). The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. Neuroscience, 96(3):451–474.

Johnson, P. M. and Kenny, P. J. (2010). Dopamine d2 receptors in addiction-like reward dysfunction and compulsive eating in obese rats. Nat Neurosci, 13(5):635–641.

Killcross, S. and Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb Cortex, 13(4):400–408.

Koob, G. (1992). Dopamine, addiction and reward. In Seminars in Neuroscience, volume 4, pages 139–148. Elsevier.

Koob, G.F. (2003). Alcoholism: allostasis and beyond. Alcohol Clin Exp Res, 27(2), 232-243

Koob, G. F. (2009). Dynamics of neuronal circuits in addiction: reward, antireward, and emotional memory. Pharmacopsychiatry, 42 Suppl 1:S32–S41.

Koob, G. F. and Moal, M. L. (2005). Plasticity of reward neurocircuitry and the 'dark side' of drug addiction. Nat Neurosci, 8(11):1442–1444.

Kravitz, A. V., Tye, L. D. and Kreitzer, A. C. (2012). Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nat Neurosci, 15, 816-818.

Laruelle, M., D D'Souza, C., Baldwin, R. M., Abi-Dargham, A., Kanes, S. J., Fingado, C. L., ... & Innis, R. B. (1997). Imaging D2 receptor occupancy by endogenous dopamine in humans. Neuropsychopharmacology, 17(3), 162-174.

Lovic, V., Saunders, B. T., Yager, L. M., and Robinson, T. E. (2011). Rats prone to attribute incentive salience to reward cues are also prone to impulsive action. Behav Brain Res, 223(2):255–261.

Maia, T. V. (2010). Two-factor theory, the actor-critic model, and conditioned avoidance. Learn Behav, 38, 50-67

Mackintosh, N. J. (1983). Conditioning and Associative Learning. Oxford University Press, Oxford, UK.

Martinez, D., Gil, R., Slifstein, M., Hwang, D. R., Huang, Y., Perez, A., ... & Abi-Dargham, A. (2005). Alcohol dependence is associated with blunted dopamine transmission in the ventral striatum. Biological psychiatry, 58(10), 779-786.

McClure, S. M., Daw, N. D., and Montague, P. R. (2003). A computational substrate for incentive salience. TINS, 26:423–8.

Meyer, P. J., Lovic, V., Saunders, B. T., Yager, L. M., Flagel, S. B., Morrow, J. D., & Robinson, T. E. (2012). Quantifying individual variation in the propensity to attribute incentive salience to reward cues. PLoS One, 7(6), e38987.

Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. J. Neurosci., 16(5):1936–47.

Morgan, D., Grant, K. A., Gage, H. D., Mach, R. H., Kaplan, J. R., Prioleau, O., Nader, S. H., Buchheimer, N., Ehrenkaufer, R. L., and Nader, M. A. (2002). Social dominance in monkeys: dopamine D2 receptors and cocaine self-administration. Nat Neurosci, 5(2):169–174.

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., and Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. Nat Neurosci, 9(8):1057–1063.

Moutoussis, M.; Bentall, R. P.; Williams, J. & Dayan, P. (2008). A temporal difference account of avoidance learning. Network, 19, 137-160

Nicola, S. M., Surmeier, D. J. and Malenka, R. C. (2000). Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. Ann. Rev. Neurosci., 23(1):185–215.

Niv, Y. (2009). Reinforcement learning in the brain. The Journal of Mathematical Psychology, 53:139–154.

Olds, J. (1956). A preliminary mapping of electrical reinforcing effects in the rat brain. J Comp Physiol Psychol, 49(3):281–285.

Ostlund, S. B. and Balleine, B. W. (2008). On habits and addiction: An associative analysis of compulsive drug seeking. Drug Discov Today Dis Models, 5(4):235–245.

Panlilio, L. V., Thorndike, E. B., and Schindler, C. W. (2007). Blocking of conditioning to a cocaine-paired stimulus: testing the hypothesis that cocaine perpetually produces a signal of larger-than-expected reward. Pharmacol Biochem Behav, 86(4):774–777.

Park, S. Q., Kahnt, T., Beck, A., Cohen, M. X., Dolan, R. J., Wrase, J., and Heinz, A. (2010). Prefrontal cortex fails to learn from reward prediction errors in alcohol dependence. J Neurosci, 30(22):7749–7753.

Parkinson, J. A., Olmstead, M. C., Burns, L. H., Robbins, T. W., and Everitt, B. J. (1999). Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by d-amphetamine. J Neurosci, 19(6):2401–2411.

Pascoli, V., Turiault, M. and Luescher C. (2012). Reversal of cocaine-evoked synaptic potentiation resets drug-induced adaptive behaviour. Nature, 481:71-75.

Phillips, P. E. M., Stuber, G. D., Heien, M. L. A. V., Wightman, R. M. and Carelli, R. M. (2003). Subsecond dopamine release promotes cocaine seeking. Nature, 422, 614-618.

Piasecki, T. M., Robertson, B. M., & Epler, A. J. (2010). Hangover and risk for alcohol use disorders: existing evidence and potential mechanisms. Current Drug Abuse Reviews, 3(2), 92-102.

Piray, P., Keramati, M. M., Dezfouli, A., Lucas, C., and Mokri, A. (2010). Individual differences in nucleus accumbens dopamine receptors predict development of addiction-like behavior: a computational approach. Neural Comput, 22(9):2334–2368.

Plessner H. Über den Begriff der Leidenschaft (1950). In: Plessner H. Condition humana. Gesammelte Schriften VIII. Frankfurt/M., Suhrkamp, 2003

Porrino, L. J., Lyons, D., Smith, H. R., Daunais, J. B., and Nader, M. A. (2004). Cocaine self-administration produces a progressive involvement of limbic, association, and sensorimotor striatal domains. J Neurosci, 24(14):3554–3562.

Puterman, M. L. (2005). Markov Decision Processes: Discrete Stochastic Dynamic Programming (Wiley Series in Probability and Statistics). Wiley-Interscience.

Redish, A. D. (2004). Addiction as a computational process gone awry. Science, 306(5703):1944–1947.

Redish, A. D., Jensen, S., and Johnson, A. (2008). A unified framework for addiction: vulnerabilities in the decision process. Behav Brain Sci, 31(4):415–37; discussion 437–87.

Robins, L. N., Davis, D. H., and Nurco, D. N. (1974). How permanent was vietnam drug addiction? Am J Public Health, 64 Suppl 12:38–43.

Richfield, E. K., Penney, J. B. and Young, A. B. (1989). Anatomical and affinity state comparisons between dopamine D1 and D2 receptors in the rat central nervous system. Neurosci., 30(3), 767-777.

Robinson, T. E. and Berridge, K. C. (2001). Incentive-sensitization and addiction. Addiction, 96(1):103–114.

Robinson, T. E. and Berridge, K. C. (2003). Addiction. Annu Rev Psychol, 54:25–53.

Robinson, T. E. and Flagel, S. B. (2009). Dissociating the predictive and incentive motivational properties of reward-related cues through the study of individual differences. Biol Psychiatry, 65(10):869–873.

Roesch, M. R., Calu, D. J., and Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci, 10(12):1615–1624.

Saxena, S., Brody, A.L., Schwartz, J.M., Baxter, L.R. (1998). Neuroimaging and frontal-subcortical circuitry in obsessive-compulsive disorder. Br J Psychiatry Suppl. (35):26-37.

Self, D. W., & Nestler, E. J. (1995). Molecular mechanisms of drug reinforcement and addiction. Ann. Rev. Neurosci, 18(1), 463-495.

Schoenbaum, G., Roesch, M. R., Stalnaker, T. A. and Takahashi, Y. K. (2009) A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. Nat Rev Neurosci, 10, 885-892.

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. Science, 275(5306):1593–1599.

Schwabe, L. and Wolf, O. T. (2009). Stress prompts habit behavior in humans. J Neurosci, 29(22):7191–7198.

Self, D. W.; Barnhart, W. J.; Lehman, D. A. and Nestler, E. J. (1996). Opposite modulation of cocaine-seeking behavior by D1- and D2-like dopamine receptor agonists. Science, 271, 1586-1589

Shallice, T. (1982). Specific impairments of planning. Philos Trans R Soc Lond B Biol Sci, 298(1089):199–209.

Shen, W., Flajolet, M., Greengard, P. and Surmeier, D. J. (2008). Dichotomous dopaminergic control of striatal synaptic plasticity. Science Signaling, 321(5890), 848.

Simon, D. A. and Daw, N. D. (2011). Neural correlates of forward planning in a spatial decision task in humans. J Neurosci, 31(14):5526–5539.

Smith, K. S., Virkud, A., Deisseroth, K. and Graybiel, A. M. (2012). Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex. Proc. Nat. Acad. Sci., 109(46), 18932-18937.

Sutton, R. (1988). Learning to predict by the methods of temporal differences. Machine Learning, 3(1):9–44.

Sutton, R. and Barto, A. (1990). Time-derivative models of Pavlovian reinforcement, pages 497–538. MIT press, Cambridge, MA.

Sutton, R. S. and Barto, A. G. (1998). Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA.

Takahashi et al.

Thanos, P. K., Volkow, N. D., Freimuth, P., Umegaki, H., Ikari, H., Roth, G., Ingram, D. K., and Hitzemann, R. (2001). Overexpression of dopamine D2 receptors reduces alcohol self-administration. J Neurochem, 78(5):1094–1103.

Tricomi, E., Balleine, B. W., and O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. Eur J Neurosci, 29(11):2225–2232.

Tsai, H.-C., Zhang, F., Adamantidis, A., Stuber, G. D., Bonci, A., de Lecea, L., and Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science, 324(5930):1080–1084.

Valentin, V. V., Dickinson, A., and O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. J Neurosci, 27(15):4019–4026.

Vanderschuren, L. J. M. J., Di Ciano, P., and Everitt, B. J. (2005). Involvement of the dorsal striatum in cue-controlled cocaine seeking. J Neurosci, 25(38):8665–8670.

Vanderschuren, L. J. M. J. and Everitt, B. J. (2004). Drug seeking becomes compulsive after prolonged cocaine self-administration. Science, 305(5686):1017–1019.

Volkow, N. D., Chang, L., Wang, G. J., Fowler, J. S., Ding, Y. S., Sedler, M., Logan, J., Franceschi, D., Gatley, J., Hitzemann, R., Gifford, A., Wong, C., and Pappas, N. (2001). Low level of brain dopamine D2 receptors in methamphetamine abusers: association with metabolism in the orbitofrontal cortex. Am J Psychiatry, 158(12):2015–2021.

Volkow, N. D., Fowler, J. S., Wang, G. J., Baler, R., and Telang, F. (2009). Imaging dopamine's role in drug abuse and addiction. Neuropharmacology, 56 Suppl 1:3–8.

Volkow, N. D., Fowler, J. S., Wang, G.-J., and Goldstein, R. Z. (2002a). Role of dopamine, the frontal cortex and

memory circuits in drug addiction: insight from imaging studies. Neurobiol Learn Mem, 78(3):610–624.

Volkow, N.D., Fowler, J.S., Wang, G.J., Hitzemann, R., Logan, J., Schlyer, D.J., Dewey, S.L., Wolf, A.P. (1993). Decreased dopamine D2 receptor availability is associated with reduced frontal metabolism in cocaine abusers. Synapse. 14(2):169-77.

Volkow, N.D., Tomasi, D., Wang, G.J., Fowler, J.S., Telang, F., Goldstein, R.Z., Alia-Klein, N., Wong, C. (2011). Reduced metabolism in brain "control networks" following cocaine-cues exposure in female cocaine abusers. PLoS One. 23;6(2):e16573.

Volkow, N. D., Wang, G. J., Fischman, M. W., Foltin, R. W., Fowler, J. S., Abumrad, N. N., Vitkun, S., Logan, J., Gatley, S. J., Pappas, N., Hitzemann, R., and Shea, C. E. (1997). Relationship between subjective effects of cocaine and dopamine transporter occupancy. Nature, 386(6627):827–830.

Volkow, N. D., Wang, G. J., Fowler, J. S., Logan, J., Hitzemann, R., Ding, Y. S., Pappas, N., Shea, C., and Piscani, K. (1996). Decreases in dopamine receptors but not in dopamine transporters in alcoholics. Alcohol Clin Exp Res, 20(9):1594–1598.

Volkow, N. D., Wang, G.-J., Fowler, J. S., Thanos, P. P. K., Logan, J., Gatley, S. J., Gifford, A., Ding, Y.-S., Wong, C., Pappas, N., and Thanos, P. (2002b). Brain DA D2 receptors predict reinforcing effects of stimulants in humans: replication study. Synapse, 46(2):79–82.

Volkow, N. D., Wang, G.-J., Telang, F., Fowler, J. S., Logan, J., Jayne, M., Ma, Y., Pradhan, K., and Wong, C. (2007). Profound decreases in dopamine release in striatum in detoxified alcoholics: possible orbitofrontal involvement. J Neurosci, 27(46):12700–12706.

Walton et al. 2010

Wassum, K. M., Ostlund, S. B., Balleine, B. W., and Maidment, N. T. (2011). Differential dependence of pavlovian incentive motivation and instrumental incentive learning processes on dopamine signaling. Learn Mem, 18(7):475–483.

Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., and Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. Psychol Sci, 22(4):490–497.

Williams, D. R. and Williams, H. (1969). Auto-maintenance in the pigeon: sustained pecking despite contingent non-reinforcement. J Exp Anal Behav, 12(4):511–520.

World Health Organization (1990). International Classification of Diseases. World Health Organization Press.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur. J. Neurosci., 19(1):181–9.