Dopamine, Reinforcement Learning, and Addiction

Author
P. Dayan

Affiliation
Gatsby Computational Neuroscience Unit, UCL, London, UK

Abstract

Dopamine is intimately linked with the modes of action of drugs of addiction. However, although its role in the initiation of drug abuse seems relatively uncomplicated, its possible involvement in the development of compulsive drug taking, and indeed vulnerability and relapse, is less clear. We first describe a modern reinforcement learning view of affective control, focusing on the roles for dopamine. We then use this as a framework to sketch various notions of the neuromodulator's possible participation in initiation and compulsion. We end with some pointers towards future theoretical developments.

Introduction

The neuromodulator dopamine is richly involved in appetitive processing. It is associated with both natural and artificial reinforcers, including intracranial self-stimulation (ICSS) and most, though not all, aspects of drugs of addiction [29, 35, 60, 68, 69, 112, 130]. However, the exact nature (and, particularly in the case of ICSS, importance [44, 55, 85]) of its involvement is incompletely understood, with theories ranging from effects on incentive motivation [99, 100, 101], and overcoming effort costs [103, 104], to learning predictions of future reward [78, 108], and beyond.

The role dopamine plays in addiction is yet more complicated, because of the very different modes of action of the different drugs, and also the rather intricate natural history of addiction, with impulsive drug taking leading to compulsive behavior and repeated cycles of withdrawal and recidivism [35, 66, 70, 113], and also with many other psychological effects such as steeper than normal discounting [120]. Further, dopamine certainly does not act alone. Rather, many other neurotransmitters and neuromodulators are involved, perhaps interacting with each other in different ways for different addictive substances. In this short note, we consider the context that reinforcement learning (RL; [118]) provides for dopamine's involvement in addiction [91, 92]. RL is a theory of adaptive, approximately-optimal, control, and offers one of the more comprehensive perspectives on the multifarious roles of dopamine in affective processing, linking computational, psychological and neurobiological notions. RL's initial treatment of dopamine was largely confined to learning, of both predictions of long run future rewards, and optimal choices of actions in the light of those predictions [7, 42, 78, 108, 116]. As pointed out by [75], this view is consonant with the notion that it is involved in the processes of 'wanting', somewhat orthogonal to the hedonic issues associated with 'liking' [10].

This RL-based account informed [91]'s suggestion about dopamine's involvement in two control-based aspects of addiction: an uncontroversial one associated with the initial reinforcing aspects of drugs, and a more contentious one [87] associated with compulsion. However, RL's scope has been extended to issues of motivation [24, 75], vigor [84], and the interactions between (i) Pavlovian and instrumental conditioning [27, 28], (ii) habitual and goal-directed behavior [20, 31], and (iii) appetitive and aversive processing [19, 25]. These models are collectively simple and rather abstract, and cannot fully reflect the many complexities of the neural substrate. However, they have proved their mettle in providing foundations for the design and understanding of a wealth of empirical approaches and results. Here, we reconsider the light that various of these new conceptions of RL sheds on dopamine's involvement in addiction.

Bibliography
DOI 10.1055/s-0028-1124107
Pharmacopsychiatry 2009;
42 (Suppl. 1): S56–S65

ISSN 0936-9528

Correspondence
P. Dayan
Gatsby Computational
Neuroscience Unit, UCL
17 Queen Square
London WC1N 3AR
dayan@gatsby.ucl.ac.uk

Section 2 very briefly sketches a modern theory of neural RL [23], pointing out the diverse influences it accords to dopamine. Section 3 considers accounts of dopamine's role in the early stages of addiction [91]. Section 4 considers possible extensions to compulsion, including the one due to [91] involving saturating predictions, one concerned with saturating action propensities [8] or boosted advantages [5], and two sorts of Pavlovian responses [10, 27]. Although we attempt to build an integrative account, it is important to remember that there are many very important differences between different drugs of addiction; further, it is not presently possible to capture all their complex effects at multiple temporal scales over the release and reception of dopamine itself and other neuromodulators, and also over other aspects of systems involved in control. Further, in keeping with the special issue, we focus particularly on dopamine, leaving many other issues associated with RL models of addiction to [92] and the extensive critical commentary associated with that paper.

Reinforcement learning and dopamine

At a computational level [73], RL offers theories of learning to predict and act appropriately in affectively charged, partially unknown, environments [118]. In the most interesting cases, the environment has multiple states (like locations in a maze), with actions causing stochastically successful transitions between states and, perhaps occasionally, giving rise to desirable or undesirable reinforcement outcomes such as foods, drugs or electric shocks. RL is often considered in instrumental or operant terms with the subject having at least partial agency. However, in the case that the subject never has a choice (which one might think of as if there is only a single action), exactly the same computational methods allow the learning of action-free predictions about future outcomes, which is normally the preserve of Pavlovian or classical conditioning. We mainly discuss RL in the richer, operant case, but refer to Pavlovian issues as they arise.

In environments such as a maze, an action cannot only be judged by its immediate consequences; rather, it is necessary to consider the cumulative utilities of all the outcomes arising in the future that depend on the action. This makes for a computationally challenging problem. RL includes different algorithmic responses to this challenge, notably a range of model-based and model-free methods [20]. In turn, these have rather different neural implementations [6], some, but not others, of which critically involve dopamine.

In model-based RL, subjects are assumed to build so-called forward models of their environments. These specify the probabilities that particular outcomes or state transitions arise from particular actions, and also report the utilities of those actions. Optimal choice in model-based RL is conceptually simple, involving forward or backwards search in the tree of all the accessible states to find the actions leading to the largest cumulative reward. It is also straightforward to handle uncertainty correctly, trading off exploration for exploitation [46, 21]. However, this conceptual simplicity is bought at what is typically a prohibitively huge computational price for searching the tree, or alternatively a prohibitively large calculational uncertainty induced by the difficulty of doing this computation accurately [20].

Since model-based decisions are made on the basis of predictions of actual outcomes, they can automatically be sensitive to the utilities of those outcomes that apply to the subject's current motivational state. In psychological terms, model-based RL is goal-directed (i.e., animals choose actions because they expect particular, desired, outcomes to result; [31]). There is evidence in rats that some aspects of goal-directed control, notably valuation, are not dependent on dopamine [32], although the expression and force of voluntary action as a whole is diminished by dopaminergic deficits [74, 83], perhaps by its effects on vigor that we discuss below.

In model-free RL, subjects acquire ways of evaluating or predicting the long-term summed utilities associated with executing actions, without building or searching in any form of forward model. Versions of these utilities include what are known as Q values [123] and advantages [5]. They can be learned in the absence of a model on the basis of the fact that predictions of long run utility should be mutually consistent along paths or trajectories. For instance, an action at one state has a high value if it leads directly to a high utility outcome, or leads to a transition to a second state that itself has a high value, or indeed both. Any inconsistency gives rise to a prediction error that can be used to correct the value of the initial state. Of course, early in learning, the value of the second state will not be accurate, and so this form of 'bootstrapping' is statistically inefficient. Nevertheless, Q and advantage values are simple to use, since they completely obviate the need for search, with actions associated with larger predicted utilities being selected more frequently.

Since the predictions are of the summed utilities of the ultimate outcomes rather the outcomes themselves, model-free control is insensitive to the current motivational state of the subject. In psychological terms, model-free RL is habitual [31]. In the end, even the relative utilities of different actions are unimportant; it is only necessary to know one that is best at each state. Thus, there is a spectrum of model-free RL methods, leading all the way down to the most ascetic architecture called the actor-critic [8]. In this, the prediction errors are used to criticize choices, ultimately enabling the learning (by the actor) of just this single best action at a state, in the absence of any information about how much better it is than other possible actions. The general impetus to perform an action at a state is sometimes called its propensity, as distinct from its Q value or advantage. We discuss this difference in more depth in section 4.

Dopamine plays a substantial role in model-free RL, in both Pavlovian [107] and instrumental [79, 102] settings, with evidence that its phasic activity [108] and release [22] represents aspects of the prediction error mentioned above. That is, it is possible rather directly to observe and measure the workings of these RL models using such techniques as electrophysiology and cyclic voltammetry, and, less directly using pharmacological functional magnetic resonance imaging (e.g., [89]). These various studies suggest that this error is coded according to the affective sign of reward (so more reward than expected leads to greater activity of dopamine neurons), and is reported by mesolimbic and mesostriatal neurons in the ventral tegmental area and substantia nigra pars compacta to targets in the amygdala, nucleus accumbens, dorsal striatum and beyond. It is believed that the phasic dopamine may influence synaptic plasticity [95, 124, 125] so as to make the predictions more accurate. However, the precise role of each of these target areas in realizing the predictions is not clear. Further, in keeping with the spectrum of increasingly austere controllers, there is a form of helical or spiralling connectivity involving a ventral-dorsal axis along the striatum and a ventral tegmental area to substantia nigra axis of the dopamine cells [50, 61, 62] providing a substrate (though

conceived slightly different there [51].) for prediction errors at a more ventral part of the spiral (acting as the critic) to teach action choices realized at a more dorsal part (the actor).

Dopamine plays at least two further roles in modern theories of RL. First, there is an association between tonic levels of the neuromodulator (which may be partially independent of phasic release; [47]) and the vigor or energy of responding [84]. This has been interpreted in RL terms as arising from the additional degree of freedom of choosing the latency of executing an action in order to balance the excess energetic cost of acting very quickly against the opportunity cost of missing out on potentially available rewards by acting very slowly. [84] suggested that tonic dopamine reports the average rate of (controllable) reward. This acts as an aspiration level - states or actions associated with rates of reward lower than the current average will be relatively aversive. In temporal terms, this average is exactly the opportunity cost of time, and, via the tradeoff mentioned above, is positively correlated with vigor. [84] discussed the account this provides of the data implicating an involvement of dopamine in effort costs [103,104]. It could also relate to the psychomotor activating properties of dopamine-boosting stimulants (which is itself a venerable idea in addiction; [131]), and increased impulsivity, since the higher the opportunity cost, the greater the price of a delay, and the less willing subjects will be to wait for rewards.

The second role concerns a phenomenon called Pavlovian to instrumental transfer (PIT). In this, subjects are separately trained on two contingencies, an instrumental one, such as lever pressing to receive one reward (A); and a Pavlovian one, of the association between a conditioned stimulus such as a tone and another reward (B). If the subject is then allowed to press the lever in extinction (i.e., without any of reward A being provided), then it will press the lever more vigorously if the tone is also played (also without the actual delivery of reward B). The greatest excess vigor comes if the Pavlovian and instrumental outcomes are literally identical (a circumstance called specific PIT); but lever pressing is enhanced even if the Pavlovian outcome is different (general PIT), provided that its current motivational value is positive (so water will not exert an effect as reward B unless the subject is thirsty). PIT appears to depend on Pavlovian values, possibly represented in the amygdala [15], affecting the nucleus accumbens, and indeed it is magnified by drugs that boost dopamine in the accumbens [133]. One idea about (general) PIT consistent with this dopaminergic influence is that presenting the Pavlovian stimulus increases the expected average reward rate, and therefore leads to enhanced vigor [84]; certainly there is a inverse correlation between the strength of activity of dopamine neurons engendered by a stimulus and the latency of the action that is inspired [105].

The final facet of modern theories of RL is an acknowledgement [13,14] of the importance of Pavlovian responses [27]. That is, model-free or model-based predictions of future appetitive (or indeed aversive) values elicit characteristic [11,12,77] preparatory and consummatory responses such as approach (or withdrawal) that can compete with, and even overwhelm [56,126], instrumental choices that experimenters impose as being necessary to get the rewards. Importantly, the nucleus accumbens, a critical site for the action of drugs of addiction, seems to be one mediator of these responses, which are organized according to a state-dependent topography over its extent [96–98]. The results on speeded responses associated with reward prediction [105] make it conceivable that dopamine influences appetitive Pavlovian responses. It is to be presumed that these responses are adaptive in ecologically relevant environments; and indeed experimenters have come to use them in working out what instrumental actions to mandate in order to hasten the course of learning in experiments. Nevertheless, instrumentally wanton actions such as approach or withdrawal in the face of reinforcers or their immediate predictors could, for instance, lead to maladaptive outcomes such as impulsivity or framing effects in choice [28].

Initiation

We first consider the dopaminergic processes that are initially engaged by drugs of addiction and that lead them, if sampled, to be likely to be repeated. One obvious route lies within model-free RL. We argued that the phasic activity of dopamine cells acts as an error associated with predictions of future reward. When this is positive, this implies that more reward than expected has been provided. In turn, dopaminergically-controlled plasticity should increase the original prediction that then appears erroneously pessimistic. If this prediction is associated with a state, then this would make the state more attractive, potentially leading to Pavlovian approach and other effects such as conditioned place preference [119]. If the prediction is the Q value or advantage of a particular action at a state, then increasing it will increase the frequency with which that action will be chosen at the state [91]. Thus, drugs that cause dopamine concentrations to be higher at key synaptic targets, by blocking reuptake, releasing it from stores, inhibiting autoreceptor-mediated feedback inhibition, or directly increasing the phasic activity of dopaminergic neurons, should lead to some of the first signs that drugs can act as (positive) reinforcers.

In the model-free actor-critic instrumental conditioning architecture, the phasic release of dopamine criticises the choice of action. In this case, drug-induced increases will inflate the propensity to perform the associated action. The apparently subtle difference between this case and the case of Q values and advantages is discussed below in terms of the evolution of compulsions.

Along with these direct routes, there are also some possibilities for indirect influence. For instance, opioids can act to magnify the effect of dopamine at its targets [30,114,115,129]. In the simplest model in which dopamine is just a pure appetitive prediction error, this would actually just change the learning rate associated with the predictions rather than acting to increase the value or propensity of states or actions (since if there is no prediction error to start with, there would be no dopamine release to be subject to opioid boost). However, there is also a baseline or tonic release of dopamine [47], and if opioids boost this, they would create an imaginary reward that would have the same effects as above [111]. There is also some evidence that novel stimuli and states lead to phasic dopamine activity [58], an effect modeled in RL prediction error terms as being a spur to exploration [64]. This could provide an initial phasic dopamine response on which the opioids would subsequently wield their effects.

Finally, there is ample evidence that opioids enhance the specific utilities of conventional outcomes such as food [9,68,88] and make various aversive outcomes less unpleasant [36]. Assuming something like a baseline level of activation of the systems involved in outcome evaluation (perhaps as part of an opponent

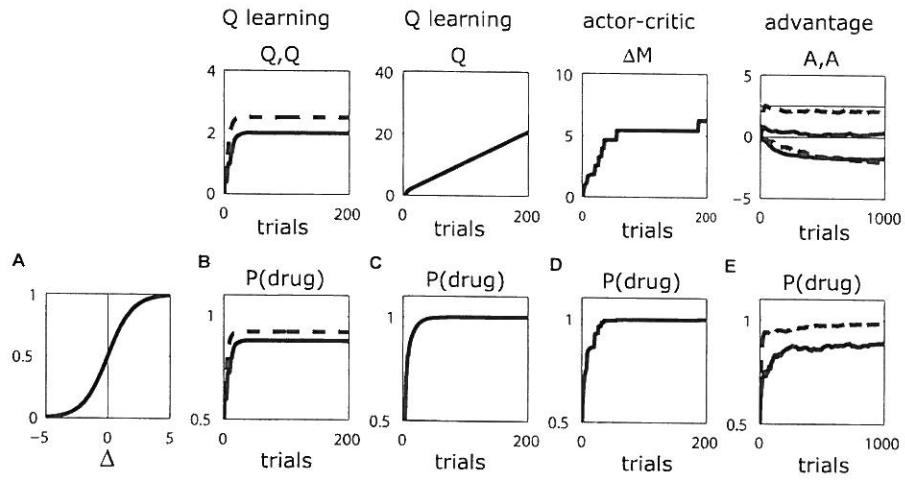

Fig. 1 Reinforcement learning action choice. A) The sigmoid function $\sigma(\Delta) = 1/(1+\exp(-\Delta))$ showing the probability of choosing the favoured action as a function of the difference in Q, propensity or advantage values. B-E) The upper plots show underlying Q values, propensities or advantages (note the different scales); the lower plots show the probabilities of choice of a target action a against a foil action b with $r_b = 0$. B) Q-learning. The upper plot shows the acquisition of a regular (solid) Q_a value for $r_a = 2$ and a drug-adjusted value (dashed) using Eq. 3 with $r_a = 0, d = 2.5$. It is apparent that this mechanism cannot lead to any unusually strong compulsion. C) Q learning with a minimum prediction error. If, as suggested by [91], the prediction error signal is always greater than d (Eq. 4), then the Q_a ultimately grows linearly. This could lead to a more deeply embedded habit, but at a cost of not accounting for data on blocking [87]. D) Actor-critic learning also leads to deeply embedded habits, with the difference in propensities between actions rising without bound. However, drugs would not exert an unusual influence here either, except perhaps increasing the speed of learning. E) In standard advantage learning (solid), the advantage of the better action tends towards 0 as the probability of choosing that action goes to 1. Given an additive drug-based factor (dashed lines), the advantage value would tend to be persistently positive, which might be one route to a compulsion.

structure; [110]) this provides an additional route by which model-free RL could be affected. Importantly, by this means, opioids, and other drugs working on specific evaluation systems, could also influence model-based RL [93]. Since the involvement of dopamine in goal-directed control is believed to be relatively constrained, this offers one route towards non-dopaminergic aspects of addiction [69,57].

Compulsion

Compared with this relatively restricted set of ideas about the earliest influences on drug-taking, the long-run behavioral and neural effects of addictive drugs appear to be more heterogeneous between different substances, and exhibit substantial variation between different individuals. They have also led to a number of rather different theoretical ideas (well aired in [92] and associated commentary). The direct relevance of RL concepts (let alone the influence of dopamine), is somewhat questionable, a fact that underlay many of the criticisms expressed to that paper. However, one of the main aspects of maladaptive decision making in addiction is the evolution of the compulsive consumption of drugs, i.e., that they are sought and consumed despite evident knowledge of their negative consequences (e.g., [33,71]). In this section, we discuss some of the key candidate RL-based routes by which dopamine's influence on at least the first stages of compulsive behavior might be explained.

Before focusing on the role that dopamine might play, we need first to understand what the structure of a compulsion might be in the context of RL. Normally, the long-run costs of actions are assumed to be weighed together with their benefits to give rise to Q or advantage values or action propensities. Given values for two such actions Q_a and Q_b (the second action could, for instance, involve just doing nothing), it is simplest [128] to consider the probability of choosing action a to be $p_a = \sigma(\Delta Q)$, where $\sigma(z) = [1/((1+\exp(-z)))]$ is the conventional logistic sigmoid, and makes a more likely the larger the value of $\Delta Q = Q_a - Q_b$. ⊙ **Fig. 1A** shows this sigmoid. One would consider a compulsion as being an unusually strongly held impetus towards an action associated with the delivery of a drug.

The first two routes to compulsion come from problems with incorrectly large action values or propensities. The alternative is that these basic action-choice quantities are correct, but that they fail to determine decision-making completely, for instance because of the malign influence of Pavlovian responses over instrumental behavior. We discuss two routes to compulsion coming from this.

It is first important to see the likely insufficiency of the simple mechanism we discussed above by which drugs that boost the dopamine signal can initiate responding. It is easiest to do this by writing down the basic RL equations (which, in this simpli-

fied case are the same as the fundamental equations of error-based learning in conditioning; [94]). Consider the case of acquiring the Q_a value of action a. If the consequence of performing action a is a reward of utility r_a, then the error associated with the current predicted value Q_a is the difference

$$\delta_a = r_a - Q_a \quad (1)$$

The suggestion is that this is reported by dopamine, giving rise to a learning rule for Q_a according to which it is changed to

$$Q_a \leftarrow Q_a + \varepsilon\delta_a \quad (2)$$

This is a discrete form of a learning rule, specifying a punctate change to Q_a (in the form of a difference equation; the arrow indicates that Q_a is replaced by a new quantity during this update) based on a single experience. It can be seen as an abstraction of a continuous time learning rule based on differential equations characterizing the (multiple timescales of) change to synapses representing this quantity. In equation 2, ε is a learning rate parameter that determines the magnitude of the change. Given a natural reward worth say $r_a = 2$ units, Eq. 2 will have Q_a approach the correct prediction of 2 exponentially quickly. This learning curve is shown in the solid line in the upper plot of ○ **Fig. 1B**; the resulting probability of choosing action a over another action b (with r_b = Q_b = 0) is shown in the lower plot.

The simplest view of the mechanism for initiation is that the drugs exert a net additive effect on the prediction error, so

$$\delta_a = r_a - Q_a + d \quad (3)$$

where $d > 0$ is a drug-induced value. In this case,[1] it is easy to see that Q_a will converge on $r_a + d$ rather than r_a. Thus, for instance, in the absence of any natural reward ($r_a = 0$), the Q value of the action will still be greater than 0 and so admit action. The dashed lines in ○ **Fig. 1B** show this case for $r_a = 0$ and $d = 2.5$. One might imagine compulsions arising if this excess reward is so great that no cost can compare. However, this does not reflect the natural history of the formation of compulsions, which appear relatively slowly, long after the values should have converged. Thus some other mechanism or mechanisms must be engaged.

Note that measuring the additional signal associated with the drugs (and indeed the other such signal discussed below) presents some challenges. Nevertheless, cyclic voltammetry seems to be becoming a method of choice (e.g., [4,16,52,90]). Findings using this method have yet to be fully integrated into modern RL accounts.

Saturating values

[91] suggested that one other factor might be an irreducible floor to the prediction error, replacing Eq. 3 with

$$\delta_a = \max\{r_a - Q_a + d, d\} \quad (4)$$

In this case, the prediction error δ_a is always greater than d. Thus, since $d > 0$, according to the learning rule in Eq. 2, the Q_a value will increase towards ∞. Of course, in reality it would have to saturate at some level, but this saturation, which might take a while to be reached, would represent a point at which balancing costs and benefits might no longer happen. ○ **Fig. 1C** shows this case, indicating the ever-increasing Q_a value associated with action leading to the drug. [91] made a critical prediction based on the difference between expressions 3 and 4. Under Eq. 3, when the prediction is correct ($Q_a = r_a + d$), there is no prediction error ($\delta_a = 0$), and so no signal that could support learning. However, under Eq. 4, there is always a signal to support learning, since $\delta_a \geq d$. The psychological phenomenon called blocking [67] probes this by including an extra stimulus and seeing if it can soak up any learning. Based on equation 4 [91], predicted that blocking should not occur for cocaine because the positive prediction error ($\delta_a > 0$), would provide the substrate for permanent learning. [87] tested this in a simple case, and found that blocking did in fact occur, weighing against this form of the learning rule.

[1] Note that this is equivalent to the first $D(S_1)$ term in Eq. 4 of [91] (P1945).

Propensities and advantages

This blocking result [87] implies that some values may be correct, and motivates the alternative proposal that action choice mechanisms associated with either propensities or advantages are disturbed instead. This is indeed one interpretation of the suggestions of [33,35], in their argument that control over the choices associated with drug seeking and taking become strongly habitual, migrating via the spiralling loops in the striatum [50,62,61] to its most dorsal extent.

Unfortunately, although there has been some interesting modeling work on the spirals [51], the rules determining which actions are controlled, and according to what learning rules, are not pinned down by the available data. However, the learning rules for both the action propensities in the actor-critic and the advantage values can be described in ways that are consistent with the spirals, and so we consider how they might be affected by drug-dependent dopamine boosts.

The actor portion of the actor-critic uses action propensities (say M_a and M_b) rather than Q values. The probability of choosing a is still $p_a = \sigma(M_a - M_b)$, but the learning rule is different. According to this, the critic learns just as in Eq. 2, except acquiring the average value V of whatever actions are tried based on the delivered reward r

$$\delta_V = r - V \quad (5)$$

$$V \leftarrow V + \varepsilon\delta_V \quad (6)$$

where, in the appetitive case, δ_V is assumed to be reported by dopamine (putatively to the ventral striatum). The blocking result [87] suggests that this proceeds normally in the face of cocaine, according to the equivalent of Eq. 3. The actor learns using the same prediction error term δ_V (but reported more dorsally, up the spiral), and changes the propensities according to which action is chosen [128]. One version of the operation of this rule is shown in ○ **Fig. 1D**, indicating the difference $\Delta M = M_a - M_b$ over trials (upper plot), together with the probability of choosing a (lower plot).

It turns out that average change to the difference between the propensities is:

$$2\eta p_a (1 - p_a)(r_a - r_b) \quad (7)$$

where η is another learning rate. Thus, if $r_a > r_b$, then ΔM increases without bound, even for natural reinforcers. This is one route to the deep embedding of a normal habit, since the greater the difference, the harder it will be to reverse the propensity. However, again, although one could imagine drugs, by boosting the dopamine signal as in Eq. 3, could boost the speed of habit formation, there is nothing in this mechanism by itself that implies that habits associated with addiction would ultimately be more deeply embedded than regular habits.

The advantage values (A_a and A_b for the two actions) are also learned using the prediction errors from value learning (Eq. 5, but by a subtly different rule. In this case, learning proceeds according to a new advantage prediction error δ_A according to

$$\delta_A = \delta_V - A_a \text{ if action a is chosen,} \quad (8)$$

changing the advantage according to

$$A_a \leftarrow A_a + \eta\delta_A \quad (9)$$

where η is a learning rate. In this case, as subjects come to choose the better action (say a) more frequently, the critic's prediction tends to the value of that action ($V \rightarrow r_a$). Thus, the prediction error for the critic tends to zero, and so, by Eq. 9, the advantage tends to zero too. The advantage of the worse action tends to be negative. The advantages and choice probabilities for $r_a = 2$, $r_b = 0$ are shown in the solid lines in the upper and lower plots of ○ **Fig. 1E**. The advantage A_a has not quite converged to 0, since the choice probability, which depends on $\sigma(A_a - A_b)$, is not 1.

If the effect of the drug on dopamine is just to increase δV by an additive factor d, as in Eq. 3, then this is just the same as a natural reinforcer with value $r_a = d$. However, if δ_A is also boosted by d, then the advantage value A_a will tend to d rather than 0, as shown by the dashed line in the upper plot of ○ **Fig. 1E**, and this could perhaps be the mark of a more deeply embedded drug habit than that associated with a natural reinforcer. To put it another way, normal reinforcers have a direct effect on value learning, and, via the spiral, an indirect effect on action learning. By manipulating the key learning signal further up the spiral, drugs that affect dopamine can have a direct effect on learned action selection too, thus leading to different, and putatively more deeply-embedded, outcomes than for normal reinforcers.

Incentive sensitization and tolerance

A different route towards explaining compulsion sees the maladaptivity as arising from the overwhelming malign influence of Pavlovian responses seen in a variety of circumstances such as omission schedules [27,126]. Indeed, this is one way to view [99,100,101]'s incentive sensitization theory of addiction (although it is perhaps not quite consistent with the authors' own view). [10]'s incentive salience theory suggests that dopamine release associated with the affective (incentive) value of stimuli makes the stimuli particularly salient and motivates the pursuit of reward (this is also related to the SEEKING notion of [86]). Compulsions arise when drugs of addiction sensitize this dopamine pathway so that the release of dopamine associated with drug-related cues leads to their capturing the whole focus of attention and forcing drug-seeking (preparatory) and drug-taking (consummatory) Pavlovian responses associated with the enhanced incentive values of the cues. The enhancement may depend, perhaps via occasion-setting [3] on the context in which the cues are presented; effects of sensitization on non-dopaminergic mechanisms may also be involved.

An important alternative to this view is that the sensitization is relative rather than absolute (see [45,132]). That is the dopamine response to conventional outcomes that are normally rewarding (and cues associated with those outcomes), may be blunted or reduced over the course of the addiction. Conversely, dopamine responses to cues associated with the addictive substance could be comparatively spared. The blunting is consistent with substantial data arguing for decreases over the course of addiction in key markers of dopamine function and action (see [53,54,121,122]), and could clearly inhibit addicted users from being tempted away by normally rewarding outcomes, leaving the addictive drug as the only target for responses. Reductions in the sensitivity of response to the drug itself, a form of systemic tolerance, has also been implicated in the apparent desire for ever-increasing doses, and this can lead to an obvious vicious cycle.

[75] reinterpret incentive salience theory in RL terms, treating incentive values of cues as predicted future rewards consequent on those cues. They note that the dopamine release associated with such cues is exactly consistent with the standard temporal difference learning model [78,117]. However [75], did not focus on the relationship between Pavlovian responses and instrumental choices, and this omission made for interpretative difficulties, since incentive salience focuses on the former, and traditional RL the latter. Newer RL notions recognize the potential contradictions between these forms of response [27,28], along with the issues such as impulsivity and framing that can come along with Pavlovian responses. The notion mentioned above that dopamine influences the appetitive Pavlovian responses allows an even closer match between RL and incentive salience.

Note the very different character of explanation arising from the notions of absolute and relative sensitization and tolerance compared with the previous notions. They consider biological neuroadaptations induced by the drugs of addiction such that the release of dopamine to cues itself changes over the course of the addiction. By contrast, the other theories consider this release to be essentially constant, and consider the effects as arising through a learning process. Of course, if the release of dopamine is indeed affected, then we might expect there not only to be Pavlovian effects (the focus of incentive sensitization), but also a range of instrumental effects, in fact on all of Q values, advantages and propensities, all pointing in the same direction of more deeply embedded drug-associated choices, which would be harder to change.

Habits versus actions

The last notion about the involvement of dopamine in compulsion is a factor that acts synergistically with incentive sensitization having to do with the interaction between model-based, goal-directed control and model-free, habitual control that we discussed above [20]. It has been suggested [27] that anomalous Pavlovian responses are more prevalent under model-free habits than model-based actions, for instance because the model-based system has the wherewithal to predict the effect of performing a maladaptive response. This, for instance, reinterprets [2]'s notion that 'the will' consists in bargaining between systems associated with short-term and long-term discounting, to one in which the task for willpower is returning control to the computationally expensive goal-directed system over the habit system, not because the latter is incorrect, or overly short term by itself, but rather because it is parasitised by maladaptive Pavlovian responses such as those discussed under incentive sensitization.

If drugs, by affecting dopamine or otherwise, can manipulate the balance in favor of habitual control (which is influenced directly)

Dayan P. Dopamine, Reinforcement Learning... Pharmacopsychiatry 2009; 42 (Suppl.1): S56–S65

over goal-directed control (which is not), then they can have the effect of weakening the will, and so enhancing the compulsive effects. Indeed, there is substantial evidence of deficits in regions of prefrontal cortex (which is associated with goal-directed control) both in animal models of addiction and human addicts themselves (see [34,122]). Theories of the balance between the systems suggest that it should be regulated in a Bayesian manner by the relative certainties of the systems rather than their relative predictions [20]. The means by which the certainties are calculated and represented is not clear (although other neuromodulators such as acetylcholine and norepinephrine, which are also influenced by various addictive drugs, may play a part; [134]). However, it could be that the predicted values are also involved in this competition, as an approximation, which would more directly implicate dopamine in the imbalance. An additional facet of incentive sensitization, namely the restriction of attention to the immediate stimuli directly associated with the drugs, could also prevent the goal-directed systems from being able to evaluate future consequences correctly, because tree search requires turning attention away from immediate stimuli to possible future states. This would leave habitual control and its Pavlovian parasites to dominate.

Discussion

We have provided a very brief review of a number of roles that the influence of drugs of addiction over dopamine might play in the initiation of drug taking and the development of compulsive behavior. We have related these in the context of a modern multifactorial theory of reinforcement learning, emphasizing the interaction between model-free and model-based systems, and between Pavlovian and instrumental conditioning. Crudely, the very early stages of drug taking are easy to accommodate, with dopamine release occasioned by drugs (or indeed other effects on specific valuation mechanisms) leading to their appearing to offer rewards. The development of compulsive behavior is more complex, with the possible involvement of various different mechanisms, only some of which may involve dopamine in any important manner. Although it is natural from the perspective of RL to look to learning as being mostly responsible [59], neuroadaptations such as those considered in incentive sensitization may also play critical roles.

In keeping with the theme of the special issue, we focused on the role of dopamine and reward processing. However, opponency between appetitive and aversive systems, which is the central idea in [70]'s hedonic homeostatic or anti-reward account of addiction, is a central focus in areas of RL [19,25,49,110], and it would be interesting to develop this link more completely. In this theory, the compulsion is a form of active avoidance behavior (which are generally hard to extinguish), controlled by the negative reinforcement associated with withdrawal, with different sorts of neuroadaptations occasioned by the drugs boosting the negative effects. There is a number of two-factor [82] RL theories of active avoidance [63,80,106] that could be adapted. It would also be worth further elaborating the theoretical ties between the effects of drugs of addiction and intra-cranial self-stimulation, particularly given the latter's also complex relationship to dopamine [44,55,85].

We have not addressed two additional facets of great importance in addiction, namely vulnerability and relapse. We can imagine various routes to vulnerabilities, notably associated with the balance between the influence of goal-directed and habitual systems (with factors favoring the latter being associated with excess vulnerability). The idea that some subjects may suffer more than others from maladaptive Pavlovian responses has recently arisen in work on sign-tracking and goal-tracking rats [37]. Sign-trackers, whose boosted Pavlovian responses (perhaps mediated by the nucleus accumbens) are evident in their excess approach to stimuli associated with reward (rather than approaching the site of the reward itself) may be more prone to the development of addiction. Indeed, in a separate study [17], showed that outbred rats with low levels of dopamine $D_{2/3}$ receptors in the nucleus accumbens were highly impulsive, and showed significantly greater escalation of cocaine intake than controls. Note also that compulsive behavior typically does not arise given only temporally limited access to drugs such as cocaine [1], perhaps because of the anti-reward effects of the repeated withdrawal.

Addiction is typically characterized by repeated cycles of abuse, abstinence and relapse [66,113]. It is hard to provide a definitive RL account of relapse, because of the complexities of the extinction processes that are presumably happening during abstinence [18,48,93]. Relapse occurring as a result of reexposure to drugs themselves or cues directly associated with drugs are relatively straightforward to accommodate; however, additional influences over relapse such as stress, and the effect of stress on the response of dopamine systems to drugs or cues [65,66,113], are more challenging. In keeping with the discussion about compulsion in section 4 about habitual mechanisms involving the dorsal striatum versus Pavlovian mechanisms involving the ventral tegmental area and the nucleus accumbens, there is evidence for both the former (e.g., [109]) and the latter (e.g., [43,76]) in relapse, with a debate concerning details of the relevant animal models of the phenomena.

In sum, RL offers a theory of control that links computational notions of optimal behavior, through the psychology of Pavlovian and operant conditioning, to the neurobiology of neuromodulators, the striatum and beyond. Thus, just as RL therefore provides a framework within which to study the impaired choices and behavior evident in psychiatric [26,72,81,127] and neurological [38,39,40,41] conditions, it should offer a window onto the maladaptive decision-making seen in addiction [91,92]. Of course, this does not mean that RL can necessarily provide an account of the teleological aspects of the development of the problems in addiction, since they may well arise through neuroadaptations that lie outside its bailiwick. The hope that the understanding that RL provides of phasic and tonic aspects of dopamine might extend to an understanding of its roles in addiction seems clearly realized for the initiation of drug taking, has some prospects for the development of compulsion, but is more of a work in progress for the critical issues of vulnerability and relapse.

Acknowledgements/Disclosure

I am very grateful to Felix Tretter for the opportunity to write this review and to him and an anonymous reviewer for their comments on the manuscript, notably the idea of relative incentive salience. Funding was from the Gatsby Charitable Foundation. I declare that I have no conflict of interest.

References

1 *Ahmed SH, Koob GF.* Transition from moderate to excessive drug intake: change in hedonic set point. Science 1998; 282 (5387): 298–300
2 *Ainslie G.* Breakdown of Will. Cambridge University Press, 3; 2001
3 *Anagnostaras SG, Schallert T, Robinson TE.* Memory processes governing amphetamine-induced psychomotor sensitization. Neuropsychopharmacology 2002; 26 (6): 703–715
4 *Aragona BJ, Cleaveland NA, Stuber GD et al.* 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 2008; 28 (35): 8821–8831
5 *Baird L.* Advantage updating. Technical Report WL-TR-93-1146. Wright-Patterson Air Force Base, OH; 1993
6 *Balleine BW.* Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. Physiol Behav 2005; 86 (5): 717–730
7 *Barto A.* Adaptive critics and the basal ganglia In Houk J, Davis J and Beiser D eds. Models of Information Processing in the Basal Ganglia. Cambridge, MA, MIT Press; 1995; 215–232
8 *Barto A, Sutton R, Anderson C.* Neuron-like adaptive elements that that can learn difficult control problems. IEEE Trans. on Systems Man and Cybernetics 1983; 13 (5): 835–846
9 *Berridge KC.* Pleasures of the brain. Brain Cogn 2003; 52 (1): 106–128
10 *Berridge KC.* The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology (Berl) 2007; 191 (3): 391–431
11 *Blanchard DC, Blanchard RJ.* Ethoexperimental approaches to the biology of emotion. Annu Rev Psychol 1988; 39: 43–68
12 *Bolles RC.* Species-specific defense reactions and avoidance learning. Psychol Rev 1970; 77: 32–48
13 *Breland K, Breland M.* The misbehavior of organisms. American Psychologist 1961; 16 (9): 681–684
14 *Breland K, Breland M.* Animal behavior. Macmillan New York; 1966
15 *Cardinal RN, Everitt BJ.* Neural and psychological mechanisms underlying appetitive learning: links to drug addiction. Curr Opin Neurobiol 2004; 14 (2): 156–162
16 *Cheer JF, Wassum KM, Heien MLAV et al.* Cannabinoids enhance subsecond dopamine release in the nucleus accumbens of awake rats. J Neurosci 2004; 24 (18): 4393–4400
17 *Dalley JW, Fryer TD, Brichard L et al.* Nucleus accumbens d2/3 receptors predict trait impulsivity and cocaine reinforcement. Science 2007; 315 (5816): 1267–1270
18 *Daw N.* Reinforcement learning models of the dopamine system and their behavioral implications. PhD thesis, Computer Science Dept. CMU; 2003
19 *Daw ND, Kakade S, Dayan P.* Opponent interactions between serotonin and dopamine. Neural Netw 2002; 15 (4–6): 603–616
20 *Daw ND, Niv Y, Dayan P.* Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 2005; 8 (12): 1704–1711
21 *Daw ND, O'Doherty JP, Dayan P et al.* Cortical substrates for exploratory decisions in humans. Nature 2006; 441 (7095): 876–879
22 *Day JJ, Roitman MF, Wightman RM et al.* Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. Nat Neurosci 2007; 10 (8): 1020–1028
23 *Dayan P.* The role of value systems in decision-making In Engel C and Singer W, eds. Better than Conscious. Ernst Strüngmann Forum, MIT Press; Cambridge MA; 2008; 51–70
24 *Dayan P, Balleine BW.* Reward, motivation, and reinforcement learning. Neuron 2002; 36 (2): 285–298
25 *Dayan P, Huys Q.* Serotonin and affective control. Annual Review of Neuroscience 2009
26 *Dayan P, Huys QJM.* Serotonin, inhibition, and negative mood. PLoS Comput Biol 2008; 4 (2): e4
27 *Dayan P, Niv Y, Seymour B et al.* The misbehavior of value and the discipline of the will. Neural Netw 2006; 19 (8): 1153–1160
28 *Dayan P, Seymour B.* Values and actions in aversion In: Glimcher P, Camerer C, Poldrack R and Fehr E, eds. Neuroeconomics: Decision making and the brain. New York, NY: Academic Press, New York, NY; 2008; 175–191
29 *Chiara G Di.* Nucleus accumbens shell and core dopamine: differential role in behavior and addiction. Behav Brain Res 2002; 137 (1–2): 75–114
30 *Chiara G Di, Imperato A.* Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats. Proc Natl Acad Sci USA 1988; 85 (14): 5274–5278
31 *Dickinson A, Balleine B.* The role of learning in motivation In: Gallistel C, ed. Stevens' handbook of experimental psychology. volume. Wiley, New York, NY, 2002; 497–5
32 *Dickinson A, Smith J, Mirenowicz J.* Dissociation of pavlovian and instrumental incentive learning under dopamine antagonists. Behav Neurosci 2000; 114 (3): 468–483
33 *Everitt BJ, Belin D, Economidou D et al.* Neural mechanisms underlying the vulnerability to develop compulsive drug-seeking habits and addiction. Philos Trans R Soc Lond B Biol Sci 2008; 363 (1507): 3125–3135
34 *Everitt BJ, Hutcheson DM, Ersche KD et al.* The orbital prefrontal cortex and drug addiction in laboratory animals and humans. Ann N Y Acad Sci 2007; 1121: 576–597
35 *Everitt BJ, Robbins TW.* Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nat Neurosci 2005; 8 (11): 1481–1489
36 *Fields HL, Heinricher MM, Mason P.* Neurotransmitters in nociceptive modulatory circuits. Annu Rev Neurosci 1991; 14: 219–245
37 *Flagel SB, Akil H, Robinson TE.* Individual differences in the attribution of incentive salience to reward-related cues: Implications for addiction. Neuropharmacology 2008
38 *Frank MJ.* Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism. J Cogn Neurosci 2005; 17 (1): 51–72
39 *Frank MJ.* Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. Neural Netw 2006; 19 (8): 1120–1136
40 *Frank MJ, Samanta J, Moustafa AA et al.* Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. Science 2007; 318 (5854): 1309–1312
41 *Frank MJ, Seeberger LC, O'reilly RC.* By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 2004; 306 (5703): 1940–1943
42 *Friston KJ, Tononi G, Reeke GN et al.* Value-dependent selection in the brain: simulation in a synthetic neural model. Neuroscience 1994; 59 (2): 229–243
43 *Fuchs RA, Evans KA, Parker MC et al.* Differential involvement of the core and shell subregions of the nucleus accumbens in conditioned cue-induced reinstatement of cocaine seeking in rats. Psychopharmacology (Berl) 2004; 176 (3–4): 459–465
44 *Gallistel CR.* The role of the dopaminergic projections in MFB self-stimulation. Behav Brain Res 1986; 20 (3): 313–321
45 *Garavan H, Pankiewicz J, Bloom A et al.* Cue-induced cocaine craving: neuroanatomical specificity for drug users and drug stimuli. Am J Psychiatry 2000; 157 (11): 1789–1798
46 *Gittins JC.* Multi-Armed Bandit Allocation Indices (Wiley Interscience Series in Systems and Optimization). John Wiley & Sons Inc., 3 1989
47 *Goto Y, Otani S, Grace AA.* The yin and yang of dopamine release: a new perspective. Neuropharmacology 2007; 53 (5): 583–587
48 *Grossberg S.* Processing of expected and unexpected events during conditioning and attention: a psychophysiological theory. Psychol Rev 1982; 89 (5): 529–572
49 *Grossberg S.* Some normal and abnormal behavioral syndromes due to transmitter gating of opponent processes. Biol Psychiatry 1984; 19 (7): 1075–1118
50 *Haber SN, Fudge JL, MacFarland NR.* Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. J Neurosci 2000; 20 (6): 2369–2382
51 *Haruno M, Kawato M.* Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. Neural Netw 2006; 19 (8): 1242–1254
52 *Heien MLAV, Khan AS, Ariansen JL et al.* Real-time measurement of dopamine fluctuations after cocaine in the brain of behaving rats. Proc Natl Acad Sci USA 2005; 102 (29): 10023–10028
53 *Heinz A, Siessmeier T, Wrase J et al.* 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 2005; 162 (8): 1515–1520
54 *Heinz A, Siessmeier T, Wrase J et al.* Correlation between dopamine d(2) receptors in the ventral striatum and central processing of alcohol cues and craving. Am J Psychiatry 2004; 161 (10): 1783–1789
55 *Hernandez G, Hamdani S, Rajabi H et al.* Prolonged rewarding stimulation of the rat medial forebrain bundle: neurochemical and behavioral consequences. Behav Neurosci 2006; 120 (4): 888–904
56 *Hershberger W.* An approach through the looking-glass. Learning & Behavior 1986; 14 (4): 443–451

57 *Hnasko TS, Sotak BN, Palmiter RD.* Morphine reward in dopamine-deficient mice. Nature 2005; 438 (7069): 854–857
58 *Horvitz JC, Stewart T, Jacobs BL.* Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. Brain Res 1997; 759 (2): 251–258
59 *Hyman SE.* Addiction: a disease of learning and memory. Am J Psychiatry 2005; 162 (8): 1414–1422
60 *Hyman SE, Malenka RC, Nestler EJ.* Neural mechanisms of addiction: the role of reward-related learning and memory. Annu Rev Neurosci 2006; 29: 565–598
61 *Ikemoto S.* Dopamine reward circuitry: two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. Brain Res Rev 2007; 56 (1): 27–78
62 *Joel D, Weiner I.* 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 2000; 96 (3): 451–474
63 *Johnson J, Li W, Li J et al.* A computational model of learned avoidance behavior in a one-way avoidance experiment. Adaptive Behavior 2002; 9 (2): 91–104
64 *Kakade S, Dayan P.* Dopamine: generalization and bonuses. Neural Netw 2002; 15 (4–6): 549–559
65 *Kalivas PW, MacFarland K.* Brain circuitry and the reinstatement of cocaine-seeking behavior. Psychopharmacology (Berl) 2003; 168 (1–2): 44–56
66 *Kalivas PW, O'Brien C.* Drug addiction as a pathology of staged neuroplasticity. Neuropsychopharmacology 2008; 33 (1): 166–180
67 *Kamin LJ.* Predictability, surprise, attention and conditioning. In Campbell BA, Church RM, eds. Punishment and aversive behavior. Appleton-Century-Crofts, New York; 1969
68 *Kelley AE, Berridge KC.* The neuroscience of natural rewards: relevance to addictive drugs. J Neurosci 2002; 22 (9): 3306–3311
69 *Koob GF.* Drugs of abuse: anatomy, pharmacology and function of reward pathways. Trends Pharmacol Sci 1992; 13 (5): 177–184
70 *Koob GF, Moal ML.* Addiction and the brain antireward system. Annu Rev Psychol 2008; 59: 29–53
71 *Koob GF, Moal ML.* Neurobiological mechanisms for opponent motivational processes in addiction. Philos Trans R Soc Lond B Biol Sci 2008; 363 (1507): 3113–3123
72 *Kumar P, Waiter G, Ahearn T et al.* Abnormal temporal difference reward-learning signals in major depression. Brain 2008; 131 (Pt 8): 2084–2093
73 *Marr D.* Vision: A computational investigation into the human representation and processing of visual information. Henry Holt and Co., Inc. New York, NY, USA; 1982
74 *Mazzoni P, Hristova A, Krakauer JW.* Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. J Neurosci 2007; 27 (27): 7105–7116
75 *MacClure SM, Daw ND, Montague PR.* A computational substrate for incentive salience. Trends Neurosci 2003; 26 (8): 423–428
76 *MacFarland K, Kalivas PW.* The circuitry mediating cocaine-induced reinstatement of drug-seeking behavior. J Neurosci 2001; 21 (21): 8655–8663
77 *MacNaughton N, Corr PJ.* A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. Neurosci Biobehav Rev 2004; 28 (3): 285–305
78 *Montague PR, Dayan P, Sejnowski TJ.* A framework for mesencephalic dopamine systems based on predictive hebbian learning. J Neurosci 1996; 16 (5): 1936–1947
79 *Morris G, Nevet A, Arkadir D et al.* Midbrain dopamine neurons encode decisions for future action. Nat Neurosci 2006; 9 (8): 1057–1063
80 *Moutoussis M, Bentall RP, Williams J et al.* A temporal difference account of avoidance learning. Network 2008; 19 (2): 137–160
81 *Moutoussis M, Williams J, Dayan P et al.* Persecutory delusions and the conditioned avoidance paradigm: towards an integration of the psychology and biology of paranoia. Cognit Neuropsychiatry 2007; 12 (6): 495–510
82 *Mowrer O.* On the dual nature of learning: A reinterpretation of conditioning and problem solving. Harvard Educational Review 1947; 17 (2): 102–150
83 *Nieoullon A, Coquerel A.* Dopamine: A key regulator to adapt action, emotion, motivation and cognition. Current Opinion in Neurology 2003; 16: S3
84 *Niv Y, Daw ND, Joel D et al.* Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology (Berl) 2007; 191 (3): 507–520
85 *Owesson-White CA, Cheer JF, Beyene M et al.* Dynamic changes in accumbens dopamine correlate with learning during intracranial self-stimulation. Proc Natl Acad Sci USA 2008; 105 (33): 11957–11962
86 *Panksepp J.* Affective Neuroscience. Oxford University Press, New York, NY; 1998
87 *Panlilio LV, Thorndike EB, Schindler CW.* 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 2007; 86 (4): 774–777
88 *Peciña S, Berridge KC.* Central enhancement of taste pleasure by intraventricular morphine. Neurobiology (Bp) 1995; 3 (3–4): 269–280
89 *Pessiglione M, Seymour B, Flandin G et al.* Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 2006; 442 (7106): 1042–1045
90 *Phillips PEM, Stuber GD, Heien MLAV et al.* Subsecond dopamine release promotes cocaine seeking. Nature 2003; 422 (6932): 614–618
91 *Redish AD.* Addiction as a computational process gone awry. Science 2004; 306 (5703): 1944–1947
92 *Redish AD, Jensen S, Johnson A.* Addiction as vulnerabilities in the decision process. Behav Brain Sci 2008; 31 (4): 461–487
93 *Redish AD, Jensen S, Johnson A et al.* Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. Psychol Rev 2007; 114 (3): 784–805
94 *Rescorla R, Wagner A.* A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory 1972; 64–99
95 *Reynolds JN, Hyland BI, Wickens JR.* A cellular mechanism of reward-related learning. Nature 2001; 413 (6851): 67–70
96 *Reynolds SM, Berridge KC.* Positive and negative motivation in nucleus accumbens shell: bivalent rostrocaudal gradients for GABA-elicited eating, taste "liking/disliking" reactions, place preference/avoidance, and fear. J Neurosci 2002; 22 (16): 7308–7320
97 *Reynolds SM, Berridge KC.* Glutamate motivational ensembles in nucleus accumbens: rostrocaudal shell gradients of fear and feeding. Eur J Neurosci 2003; 17 (10): 2187–2200
98 *Reynolds SM, Berridge KC.* Emotional environments retune the valence of appetitive versus fearful functions in nucleus accumbens. Nat Neurosci 2008; 11 (4): 423–425
99 *Robinson TE, Berridge KC.* Incentive-sensitization and addiction. Addiction 2001; 96 (1): 103–114
100 *Robinson TE, Berridge KC.* Addiction. Annu Rev Psychol 2003; 54: 25–53
101 *Robinson TE, Berridge KC.* The incentive sensitization theory of addiction: some current issues. Philos Trans R Soc Lond B Biol Sci 2008; 363 (1507): 3137–3146
102 *Roesch MR, Calu DJ, Schoenbaum G.* Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci 2007; 10 (12): 1615–1624
103 *Salamone JD, Correa M, Farrar A et al.* Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. Psychopharmacology (Berl) 2007; 191 (3): 461–482
104 *Salamone JD, Correa M, Mingote S et al.* Nucleus accumbens dopamine and the regulation of effort in food-seeking behavior: implications for studies of natural motivation, psychiatry, and drug abuse. J Pharmacol Exp Ther 2003; 305 (1): 1–8
105 *Satoh T, Nakai S, Sato T et al.* Correlated coding of motivation and outcome of decision by dopamine neurons. J Neurosci 2003; 23 (30): 9913–9923
106 *Schmajuk N, Zanutto B.* Escape, avoidance, and imitation: A neural network approach. Adaptive Behavior 1997; 6 (1): 63
107 *Schultz W.* Getting formal with dopamine and reward. Neuron 2002; 36 (2): 241–263
108 *Schultz W, Dayan P, Montague PR.* A neural substrate of prediction and reward. Science 1997; 275 (5306): 1593–1599
109 *See RE, Elliott JC, Feltenstein MW.* The role of dorsal vs. ventral striatal pathways in cocaine-seeking behavior after prolonged abstinence in rats. Psychopharmacology (Berl) 2007; 194 (3): 321–331
110 *Solomon RL, Corbit JD.* An opponent-process theory of motivation. I. temporal dynamics of affect. Psychol Rev 1974; 81 (2): 119–145
111 *Spanagel R, Herz A, Shippenberg TS.* Opposing tonically active endogenous opioid systems modulate the mesolimbic dopaminergic pathway. Proc Natl Acad Sci USA 1992; 89 (6): 2046–2050
112 *Spanagel R, Weiss F.* The dopamine hypothesis of reward: past and current status. Trends Neurosci 1999; 22 (11): 521–527

113 *Stewart J.* Psychological and neural mechanisms of relapse. Philos Trans R Soc Lond B Biol Sci 2008; 363 (1507): 3147–3158
114 *Stinus L, Cador M, Moal ML.* Interaction between endogenous opioids and dopamine within the nucleus accumbens. Ann N Y Acad Sci 1992; 654: 254–273
115 *Stinus L, Koob GF, Ling N et al.* Locomotor activation induced by infusion of endorphins into the ventral tegmental area: evidence for opiate-dopamine interactions. Proc Natl Acad Sci USA 1980; 77 (4): 2323–2327
116 *Suri RE, Schultz W.* A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. Neuroscience 1999; 91 (3): 871–890
117 *Sutton R.* Learning to predict by the methods of temporal differences. Machine Learning 1988; 3 (1): 9–44
118 *Sutton RS, Barto AG.* Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning). The MIT Press, 3 1998
119 *Tzschentke TM.* Measuring reward with the conditioned place preference paradigm: a comprehensive review of drug effects, recent progress and new issues. Prog Neurobiol 1998; 56 (6): 613–672
120 *Verdejo-Garcia A, Lawrence AJ, Clark L.* Impulsivity as a vulnerability marker for substance-use disorders: review of findings from high-risk research, problem gamblers and genetic association studies. Neurosci Biobehav Rev 2008; 32 (4): 777–810
121 *Volkow ND, Fowler JS, Wang G-J et al.* Dopamine in drug abuse and addiction: results from imaging studies and treatment implications. Mol Psychiatry 2004; 9 (6): 557–569
122 *Volkow ND, Fowler JS, Wang G-J et al.* Dopamine in drug abuse and addiction: results of imaging studies and treatment implications. Arch Neurol 2007; 64 (11): 1575–1579
123 *Watkins C.* Learning from Delayed Rewards. PhD thesis, University of Cambridge; 1989
124 *Wickens J.* Striatal dopamine in motor activation and reward-mediated learning: steps towards a unifying model. J Neural Transm Gen Sect 1990; 80 (1): 9–31
125 *Wickens JR, Reynolds JNJ, Hyland BI.* Neural mechanisms of reward-related motor learning. Curr Opin Neurobiol 2003; 13 (6): 685–690
126 *Williams DR, Williams H.* Auto-maintenance in the pigeon: sustained pecking despite contingent non-reinforcement. J Exp Anal Behav 1969; 12 (4): 511–520
127 *Williams J, Dayan P.* Dopamine, learning, and impulsivity: a biological account of attention-deficit/hyperactivity disorder. J Child Adolesc Psychopharmacol 2005; 15 (2): 160–179; discussion 157–159
128 *Williams R.* Simple statistical gradient-following algorithms for connectionist reinforcement learning. Reinforcement Learning 1992; 8: 229–256
129 *Wise RA.* Opiate reward: sites and substrates. Neurosci Biobehav Rev 1989; 13 (2–3): 129–133
130 *Wise RA.* Forebrain substrates of reward and motivation. J Comp Neurol 2005; 493 (1): 115–121
131 *Wise RA, Bozarth MA.* A psychomotor stimulant theory of addiction. Psychol Rev 1987; 94 (4): 469–492
132 *Wrase J, Schlagenhauf F, Kienast T et al.* Dysfunction of reward processing correlates with alcohol craving in detoxified alcoholics. Neuroimage 2007; 35 (2): 787–794
133 *Wyvell CL, Berridge KC.* Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: enhancement of reward "wanting" without enhanced "liking" or response reinforcement. J Neurosci 2000; 20 (21): 8122–8130
134 *Yu AJ, Dayan P.* Uncertainty, neuromodulation, and attention. Neuron 2005; 46 (4): 681–692