REVIEW ARTICLE
Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks

Henry H. Yin,[1] Sean B. Ostlund[2] and Bernard W. Balleine[2]

[1]Department of Psychology and Neuroscience, Center for Cognitive Neuroscience, Duke University, 103 Research Drive, Box 91050, Durham, NC 27708, USA
[2]Department of Psychology and Brain Research Institute, University of California, Los Angeles, CA, USA

Abstract

Here we challenge the view that reward-guided learning is solely controlled by the mesoaccumbens pathway arising from dopaminergic neurons in the ventral tegmental area and projecting to the nucleus accumbens. This widely accepted view assumes that reward is a monolithic concept, but recent work has suggested otherwise. It now appears that, in reward-guided learning, the functions of ventral and dorsal striata, and the cortico-basal ganglia circuitry associated with them, can be dissociated. Whereas the nucleus accumbens is necessary for the acquisition and expression of certain appetitive Pavlovian responses and contributes to the motivational control of instrumental performance, the dorsal striatum is necessary for the acquisition and expression of instrumental actions. Such findings suggest the existence of multiple independent yet interacting functional systems that are implemented in iterating and hierarchically organized cortico-basal ganglia networks engaged in appetitive behaviors ranging from Pavlovian approach responses to goal-directed instrumental actions controlled by action–outcome contingencies.

Introduction

It has become common in the recent literature to find a monolithic concept of 'reward' applied uniformly to appetitive behavior, whether to denote anything that is good for the organism (usually from the perspective of the experimenter), or used interchangeably with older terms like 'reinforcement' or 'incentive'. This state of affairs is encouraged by, if not itself the consequence of, the focus on a single neural substrate for 'reward' involving release of dopamine (DA) in the nucleus accumbens (Berke & Hyman, 2000; Grace *et al.*, 2007).

The link between the mesoaccumbens pathway and reward, recognized decades ago, has been reinvigorated by more recent evidence that the phasic DA signal encodes a reward prediction error, which presumably serves as a teaching signal in associative learning (Schultz *et al.*, 1997). According to the most popular interpretation, just as there is a single signal for reward, so there is a single signal for reward-guided learning, which in this case means association between a stimulus and a reward (Montague *et al.*, 2004). The question of how this type of learning controls adaptive behavior has, however, been neglected; it is simply assumed that the DA signal is sufficient for both predictive learning and the conditional responses (CR) engendered thereby, and for goal-directed actions guided by their association with reward. Consequently, the focus of most research in the field of reward and addiction is DA signaling and related plasticity in the mesoac-

cumbens pathway (Berridge & Robinson, 1998; Hyman *et al.*, 2006; Grace *et al.*, 2007).

This view of the reward process, as is increasingly recognized (Cardinal *et al.*, 2002; Balleine, 2005; Everitt & Robbins, 2005; Hyman *et al.*, 2006), is both inadequate and misleading. It is inadequate because neither the acquisition nor the performance of goal-directed actions can be explained in terms of the associative processes that mediate stimulus–reward learning. It is misleading, moreover, because the exclusive focus on activity in the mesoaccumbens pathway, which is neither necessary nor sufficient for goal-directed actions, has diverted attention from the more fundamental question of exactly what goal-directed actions are and how they are implemented by the brain. Indeed, according to converging evidence from a variety of experimental approaches, what has previously appeared to be a single reward mechanism may in fact comprise multiple processes with distinct behavioral effects and neural substrates (Corbit *et al.*, 2001; O'Doherty *et al.*, 2004; Yin *et al.*, 2004, 2005b; Delgado *et al.*, 2005; Haruno & Kawato, 2006a; Tobler *et al.*, 2006, 2007; Jedynak *et al.*, 2007; Robinson *et al.*, 2007).

Here we attempt to expose some of the problems associated with the current mesoaccumbens model and to propose, in its place, a different model of reward-guided learning. We shall argue that the striatum is a highly heterogeneous structure that can be divided into at least four functional domains, each of which acts as a hub in a distinct functional network with other cortical, thalamic, pallidal and midbrain components. The integrative functions of these networks, ranging from the production of unconditional responses (UR) elicited by reward to the

Correspondence: Dr H. H. Yin, as above.
E-mail: hy43@duke.edu

control of goal-directed actions, can be dissociated and studied using contemporary behavioral assays.

Prediction and control

The mesoaccumbens pathway is often assumed to be necessary for the acquisition of an association between reward and environmental stimuli that predict that reward. For example, in some of the experiments examining the phasic activity of DA cells elicited by reward, monkeys were trained to associate a stimulus with the delivery of juice (Waelti *et al.*, 2001) and subsequently respond to the stimulus with a CR – anticipatory licking. The monkey's licking could be goal-directed, because it believes it is necessary to obtain juice. Alternatively, licking can be elicited by the antecedent stimulus with which juice is associated. Which of these determinants of the monkeys' licking is controlling the behavior in any particular situation is not known *a priori*, and cannot be determined by superficial observation; it can only be determined using tests designed specifically for this purpose. These tests, which have taken many decades to develop, form the core of the major modern advances in the study of learning and behavior (Table 1). From the use of these tests, to be discussed below, we now know that the same behavioral response – whether it is ambulatory approach, orienting or pressing a lever – can arise from multiple influences that are experimentally dissociable.

Insensitivity to the central ambiguity in the actual determinants of behavior is thus the chief problem with current neuroscientific analysis of reward-guided learning. To understand the significance of this problem, it is necessary to appreciate the differences between how predictive (or Pavlovian) learning and goal-directed (or instrumental) learning control appetitive behavior. Indeed, judging by how often these two processes have been conflated in the literature on reward, a brief review of this distinction seems to be a useful starting point for our discussion.

In appetitive Pavlovian conditioning, the reward (i.e. the unconditional stimulus or US) is paired with a stimulus (conditional stimulus or CS), regardless of the animal's behavior, whereas in instrumental learning the reward is contingent upon the animals' actions. The critical question in both situations is, however, whether the stimulus–reward association or the action–reward association is controlling behavior. As simple as it seems, this question eluded investigators for many decades largely because the behavioral responses in these situations can appear identical. Thus, the CRs controlled by the Pavlovian stimulus–reward association can often have a veneer of goal-directedness about them. Even salivation, Pavlov's original CR, could have been produced by his dogs as a deliberate attempt to facilitate ingestion. It is precisely because of this ambiguity that the most obvious explanation – namely that in Pavlovian conditioning the stimulus–outcome association is learned, whereas in instrumental conditioning the action–outcome association is learned – failed to garner much support for many decades (Skinner, 1938; Ashby, 1960; Bolles, 1972; Mackintosh, 1974). Nevertheless, although many Pavlovian CRs are autonomic or consummatory, other CRs, such as approach behavior towards a reward, are not so conveniently characterized (Rescorla & Solomon, 1967); indeed, they can easily be mistaken for instrumental actions (Brown & Jenkins, 1968; Williams & Williams, 1969; Schwartz & Gamzu, 1977). We now know that, despite a superficial resemblance, Pavlovian CRs and goal-directed instrumental actions differ in the representational structure controlling performance of the response (Schwartz & Gamzu, 1977).

The most direct means of establishing whether the performance of a response is mediated by a stimulus–reward or an action–reward association is to examine the specific contingency controlling performance. The example of salivation is instructive here. Sheffield (1965) tested whether salivation in Pavlovian conditioning was controlled by its relationship to reward or by the stimulus–reward association. In his experiment, dogs received pairings between a tone and a food reward (Sheffield, 1965). However, if the dogs salivated during the tone, then the food was not delivered on that trial. This arrangement maintained a Pavlovian relationship between the tone and food, but abolished any direct association between salivation and food delivery. If the salivation was an action controlled by its relationship to food, then the dogs should stop salivating – indeed they should never acquire salivation to the tone at all. Sheffield found that it was clearly the Pavlovian tone–food relationship that controlled the salivation CR. During the course of over 800 tone–food pairings, the dogs acquired and maintained salivation to the tone even though this resulted in their losing most of the food they could have obtained by not salivating. A similar conclusion was reached by others in studies with humans (Pithers, 1985) and other animals (Brown & Jenkins, 1968; Williams & Williams, 1969; Holland, 1979); in all cases it appears that, despite their great variety, Pavlovian responses are not controlled by their relationship to the reward – i.e. by the action–outcome contingency.

The term contingency refers to the conditional relationship between an event 'A' and another, 'B', such that the occurrence of B depends on A. A relationship of this kind can readily be degraded by presenting B in the absence of A. This experimental manipulation, referred to as contingency degradation, is commonly performed by presenting a reward independently of either the predictive stimulus or the action. Although this approach was originally developed to study Pavlovian conditioning (Rescorla, 1968), instrumental contingency degradation has also become a common tool (Hammond, 1980). When these

TABLE 1. A classification of reward-guided learning and behavior based on the type of responses being modified by experience

	Pavlovian learning		Instrumental learning	
	Consummatory responses, e.g. orofacial responses to taste	Preparatory responses, e.g. anticipatory approach: S–O	Goal-directed actions: A–O	Stimulus-driven habits: S–R
Sensitive to:				
Outcome devaluation?	Yes	Yes	Yes	No
Changes in S–O contingency?	Yes	Yes	No	No?
Changes in A–O contingency?	No	No	Yes	No?

The content of learning can be experimentally probed using contemporary behavioral assays. For entries with question marks, no data are currently available. A–O, action–outcome; S–O, stimulus–outcome; S–R, stimulus–response.

contingencies are directly manipulated, the content of learning is revealed: e.g. in autoshaping, a Pavlovian CR 'disguised' as an instrumental action is disrupted by manipulations of the Pavlovian rather than the instrumental contingency (Schwartz & Gamzu, 1977).

Goal-directed instrumental actions are characterized by two criteria: (1) sensitivity to changes in the value of the outcome; and (2) sensitivity to changes in the contingency between action and outcome (Dickinson, 1985; Dickinson & Balleine, 1993). Sensitivity to outcome devaluation alone, it should be emphasized, does not suffice in characterizing a response as goal-directed because some Pavlovian responses can also be sensitive to this manipulation (Holland & Rescorla, 1975). The performance of goal-directed instrumental actions is also sensitive to manipulations of the action–outcome contingency, whereas Pavlovian responses are sensitive to manipulations of the stimulus–outcome contingency (Rescorla, 1968; Davis & Bitterman, 1971; Dickinson & Charnock, 1985). An important exception, however, can be found in the case of habits (see below), which are more similar to Pavlovian responses in their relative insensitivity to changes in the instrumental contingency, but are also impervious to outcome devaluation because the outcome is not part of the representational structure controlling performance (cf. Dickinson, 1985; and below for further discussion).

To summarize then, it is of the utmost importance that a particular response be clearly defined in terms of the controlling contingency rather than by either the response form or the behavioral task used to establish it. Without examining the controlling contingency in a given situation, both the behavior and the neural processes found to mediate the behavior are likely to be mischaracterized. Ultimately, as we shall argue, it is the actual contingencies, acquired through learning and implemented by distinct neural systems, that control behavior, though they may share the same 'final common pathway'. Thus, the central challenge is to go beyond appearances to uncover the underlying contingency controlling behavior (for a summary, see Table 1). In order to claim that specific neural structures mediate specific psychological capacities, e.g. goal-directedness, the status of the behavior must be assessed with the appropriate behavioral assays. To do otherwise is to invite confusion as groups argue over the appropriate neural determinants whilst failing to recognize that their behavioral tasks could be measuring different phenomena. What matters, ultimately, is what the animal actually learns, not what the experimenter believes that the animal learns, and what the animal actually learns can only be revealed by assays that directly probe the content of learning.

The Pavlovian-instrumental distinction would have been trivial if the animal managed to learn the same thing (say an association between the stimulus and reward) no matter what the experimental arrangements are. Using the most common measures of learning available to neuroscience today, there is simply no way to tell. Thus, researchers often claim to study goal-directed behavior without examining whether the behavior in question is in fact directed towards the goal. Although different types of learning are commonly assumed to result from the use of different 'tasks' or 'paradigms', more often than not researchers fail to provide an adequate rationale for their assumptions.

A classic example of this issue is the use of mazes to study learning. One problem with maze experiments and related assays, like conditioned place preference, is the difficulty of experimentally dissociating the influence of the Pavlovian (stimulus–reward) and the instrumental (action–reward) contingencies on behavior (Dickinson, 1994; Yin & Knowlton, 2002). Thus, moving through a T-maze to get food could reflect a response strategy (turn left) or simply a conditioned approach towards some extra-maze landmark controlled

by the cue–food association (Restle, 1957). One way of testing whether the latter plays a role in performance is to invert the maze; now response learners should continue to turn left whereas those using extra-maze cues should turn right. But are those that continue to turn left really using a response strategy or are they merely approaching some intra-maze cue associated with food? It is not a simple matter to find out, because the usual controls for Pavlovian control of behavior cannot easily be applied in maze studies. One of these, the bidirectional control, establishes that animals can exert control over a particular response by requiring the reversal of the direction of that response to earn reward (Hershberger, 1986; Heyes & Dawson, 1990). Unfortunately, in a maze, response reversal may still not be sufficient to establish an action as goal-directed, because reversal can be accomplished by extinguishing the existing stimulus–reward relationship and substituting it with another. For example, a rat approaching a particular intra-maze cue may learn, during reversal, that it is no longer paired with reward, but that some other stimulus is, resulting in acquiring an approach CR towards the new stimulus. Thus, they can apparently reverse their response without having ever encoded the response–reward contingency. Because this possibility cannot be tested in practice, the use of mazes, place preference procedures or simple locomotor tasks to study goal-directed actions is particularly perilous and likely to result in mischaracterizing the processes controlling behavior together with the specific role of any neural processes found to be involved (Smith-Roe & Kelley, 2000; Hernandez *et al.*, 2002; Atallah *et al.*, 2007).

Nucleus accumbens is not necessary for instrumental learning

The inadequacies of current behavioral analysis become particularly clear in the study of the nucleus accumbens. Many studies have suggested that this structure is critical for the acquisition of goal-directed actions (Hernandez *et al.*, 2002, 2005; Goto & Grace, 2005; Pothuizen *et al.*, 2005; Taha & Fields, 2006; Atallah *et al.*, 2007; Cheer *et al.*, 2007; Lerchner *et al.*, 2007). But this conclusion has been reached based largely on measures of a change in performance alone, using tasks in which the contingency controlling behavior is ambiguous. Although the observation that a manipulation impairs the acquisition of some behavioral response could indicate a learning deficit, they could also reflect an effect on response initiation or motivation. For example, an impairment in the acquisition of lever pressing can often reflect an effect on performance rather than on learning (Smith-Roe & Kelley, 2000). Acquisition curves alone, as incomplete representations of any learning process, must be interpreted with caution (Gallistel *et al.*, 2004). Unfortunately, the distinction between learning and performance, perhaps the oldest lesson in the study of learning, is often ignored today.

A more detailed analysis indicates that the accumbens is neither necessary nor sufficient for instrumental learning. Lesions of the accumbens shell do not alter sensitivity of performance to outcome devaluation (de Borchgrave *et al.*, 2002; Corbit *et al.*, 2001) or to instrumental contingency degradation (Corbit *et al.*, 2001), whereas lesions of the accumbens core have been found to reduce sensitivity to devaluation without impairing the rats' sensitivity to selective degradation of the instrumental contingency (Corbit *et al.*, 2001). Other studies assessing the effect of accumbens manipulations on the acquisition of a new response in studies of conditioned reinforcement have consistently found an effect on reward-related performance, particularly the enhancement of performance by amphetamine, but not on the acquisition of responding *per se* (Parkinson *et al.*, 1999).

Likewise, a systematic study by Cardinal and Cheung also found no effect of accumbens core lesions on acquisition of a lever press response under a continuous reinforcement schedule; impaired acquisition was only observed with delayed reinforcement (Cardinal & Cheung, 2005).

Although the accumbens does not encode the instrumental contingency (Corbit *et al.*, 2001), considerable evidence suggests that it does play a fundamental role in instrumental 'performance', a role that we can now better define in light of recent work. As concluded by several studies, the accumbens is critical for certain types of appetitive Pavlovian conditioning, and mediates both the non-specific excitatory effects that reward-associated cues can have on instrumental performance, as well as the outcome-specific biases on response selection produced by such cues. Lesions of the core, or of the anterior cingulate, a major source of cortical input to the core, or a disconnection between these two structures impair the acquisition of Pavlovian approach behavior (Parkinson *et al.*, 2000). Local infusion of a D1-like DA receptor antagonist or a *N*-methyl-D-aspartate (NMDA) glutamate receptor antagonist immediately after training also impaired this form of learning without affecting performance (Dalley *et al.*, 2005). These data agree with measures of *in vivo* neural activity. For example, Carelli and colleagues found that neurons in the accumbens core can change their activity systematically during the learning of a Pavlovian autoshaping task (Day *et al.*, 2006; Day & Carelli, 2007).

Neurons in the shell region appear to be tuned to rewards and aversive stimuli, even before any learning experience; they are also capable of developing responses to CSs that predict these outcomes (Roitman *et al.*, 2005). Work by Berridge and colleagues, moreover, has raised the possibility that certain regions within the nucleus accumbens shell and in the downstream ventral pallidum may be characterized as 'hedonic hotspots'. These areas directly modulate unconditional hedonic responses to rewards, such as taste reactivity. For example, agonists of opioid receptors in these regions can significantly amplify ingestive taste reactivity to sucrose. Such highly localized regions, however, are embedded in wider networks that do not play a role in consummatory appetitive behavior (Taha & Fields, 2005, 2006; Pecina *et al.*, 2006).

The distinction in the relative roles of core and shell appears to be one between preparatory and consummatory appetitive behaviors, respectively, which can be easily modified by experience through distinct types of Pavlovian conditioning. Preparatory responses such as approach are linked with general emotional qualities of the outcome, whereas the consummatory behaviors are linked with more specific sensory qualities; they are also differentially susceptible to different types of CS, e.g. preparatory responses are more readily conditioned using a stimulus with a long duration (Konorski, 1967; Dickinson & Dearing, 1979; Balleine, 2001; Dickinson & Balleine, 2002).

At any rate, the evidence implicating the accumbens in some aspects of Pavlovian conditioning is overwhelming. It is, however, not the only structure involved, and other networks, such as those involving the various amygdaloid nuclei, also appear to play a central role in both the preparatory and consummatory components of Pavlovian conditioning (Balleine & Killcross, 2006).

One function that can clearly be attributed to the accumbens is the integration of Pavlovian influences on instrumental behavior. Pavlovian CRs, including those reflecting the activation of central motivational states, such as craving and arousal, can exert a strong influence on the performance of instrumental actions (Trapold & Overmier, 1972; Lovibond, 1983; Holland, 2004). For instance, a CS that independently predicts food delivery can increase instrumental responding for the very same food. This effect is commonly studied using the Pavlovian-instrumental transfer paradigm (PIT). In PIT, animals receive separate Pavlovian and instrumental training phases, in which they learn, independently, to associate a cue with food and to press a lever for the same food. Then on probe trials, the cue is presented with the lever available, and the elevation of response rates in the presence of the CS is measured. Two forms of PIT have been identified; one related to the generally arousing effect of reward-related cues and a second more selective effect on choice performance produced by the predictive status of a cue with respect to one specific reward as opposed to others. The accumbens shell is necessary for this latter outcome-specific form of PIT, but is neither necessary for the former, more general form nor for sensitivity to outcome devaluation; by contrast, lesions of the accumbens core reduce sensitivity to both outcome devaluation and the general form of PIT, but leave intact outcome-specific PIT (Corbit *et al.*, 2001).

A recent study provided further insight into the role of the accumbens shell in outcome-specific PIT (Wiltgen *et al.*, 2007). Controlled expression of active calcium/calmodulin-dependent protein kinase II (CaMKII) in the striatum did not affect instrumental or Pavlovian learning, but abolished specific PIT. This deficit in PIT was not permanent and could be reversed by turning off the transgene expression with doxycycline, demonstrating that the deficit was associated with performance only. Artificially enhancing the level of CaMKII in the striatum therefore blocks the outcome-specific transfer of incentive motivation from the Pavlovian to the instrumental system. Interestingly, turning on the CaMKII transgene was also found to reduce the excitability of neurons in the accumbens shell, without affecting basal transmission or synaptic strength.

The dorsal striatum

The dorsal striatum, also known as the neostriatum or caudate-putamen, receives massive projections from the so-called neocortex. It can be further divided into an associative region, which in rodents is more medial and continuous with the ventral striatum, and a sensorimotor region, which is more lateral (Groenewegen *et al.*, 1990; Joel & Weiner, 1994). As a whole, the dorsal striatum is innervated by DA cells from the substantia nigra pars compacta (SNc), and only receives meager projections from the ventral tegmental area (VTA) DA neurons (Joel & Weiner, 2000). Previous work on the dorsal striatum has focused mostly on its role in stimulus–response (S–R) habit learning (Miller, 1981; White, 1989). This view is based on the law of effect, according to which a reward acts to strengthen, or reinforce, an S–R association between the environmental stimuli and the response performed, as a result of which the tendency to perform that response increases in the presence of those stimuli (Thorndike, 1911; Hull, 1943; Miller, 1981). Thus, the corticostriatal pathway is thought to mediate S–R learning, with DA acting as the reinforcement signal (Miller, 1981; Reynolds & Wickens, 2002).

S–R models have the advantage of containing a parsimonious rule for translating learning into performance. A model based on action-related expectancies, by contrast, is more complicated because the belief 'Action A leads to Outcome O' does not necessarily have to be translated into action (Guthrie, 1935; Mackintosh, 1974); information of this kind can be used both to perform 'A' and to avoid performing 'A'. For this reason, traditional theories shunned the most obvious explanation – namely that animals can acquire an action–outcome contingency that guides choice behavior. The last few decades, however, have seen a substantial revision of the law of effect (Adams, 1982; Colwill & Rescorla, 1986; Dickinson, 1994; Dickinson *et al.*,

1996). The results of many studies have demonstrated that instrumental actions can be truly goal-directed, i.e. sensitive to changes in reward value as well as the causal efficacy of the action (see Dickinson & Balleine, 1993, 2002; Balleine, 2001 for reviews). Nevertheless, over the course of extensive training under constant conditions, even newly acquired actions can become relatively automatic and stimulus-driven – a process known as habit formation (Adams & Dickinson, 1981; Adams, 1982; Yin *et al.*, 2004). Habits thus defined, being automatically elicited by antecedent stimuli, are not controlled by the expectancy or representation of the outcome, they are consequently impervious to changes in outcome value. From this perspective, the law of effect is therefore a special case that applies only to habitual behavior.

The current classification of instrumental behavior divides it into two classes. The first class comprises goal-directed actions controlled by the instrumental contingency; the second, habitual behavior impervious to changes in outcome value (Table 1). Using behavioral assays like outcome devaluation and instrumental contingency degradation, Yin *et al.* established a functional dissociation between the sensorimotor (dorsolateral striatum, DLS) and associative regions (dorsomedial striatum, DMS) of the dorsal striatum (Yin & Knowlton, 2004; Yin *et al.*, 2004, 2005a,b, 2006a). Lesions of the DLS impaired the development of habits, resulting in a more goal-directed mode of behavioral control. Lesions of the DMS have the opposite effect and result in a switch from goal-directed to habitual control. Yin *et al.* concluded, therefore, that the DLS and DMS can be functionally dissociated in terms of the type of associative structures they support: the DLS is critical for habit formation, whereas the DMS is critical for acquisition and expression of goal-directed actions. This analysis predicts that, under certain conditions (e.g. extended training), the control of actions can shift from the DMS-dependent system to the DLS-dependent system, a conclusion that is in broad agreement with the considerable literature on primates, including human neuroimaging (Hikosaka *et al.*, 1989; Jueptner *et al.*, 1997a; Miyachi *et al.*, 1997, 2002; Delgado *et al.*, 2004, 2005; Haruno *et al.*, 2004; Tricomi *et al.*, 2004; Samejima *et al.*, 2005; Haruno & Kawato, 2006a,b; Lohrenz *et al.*, 2007; Tobler *et al.*, 2007). It should be remembered, of course, that physical location (e.g. dorsal or ventral) alone cannot be a reliable guide in comparing the rodent striatum and the primate striatum, such comparisons should be made with caution, after careful consideration of the anatomical connectivity.

The effects of dorsal striatal lesions can be compared with those of accumbens lesions (Smith-Roe & Kelley, 2000; Atallah *et al.*, 2007). As already mentioned, the standard tests for establishing a behavior as 'goal-directed' are outcome devaluation and degradation of the action–outcome contingency (Dickinson & Balleine, 1993). Lesions of the DMS render behavior insensitive to both manipulations (Yin *et al.*, 2005b), whereas lesions of the accumbens core or shell do not (Corbit *et al.*, 2001). Moreover, the probe tests of these behavioral assays are typically conducted in extinction, without the presentation of any reward, in order to assess what the animal has learned without contamination by new learning. They thus directly probe the representational structure controlling behavior. As an additional experimental control, it is often useful to conduct a separate devaluation test in which rewards are actually delivered – the so-called 'rewarded test.' Lesions of the DMS did not abolish sensitivity to outcome devaluation on the rewarded test, as should be expected as the delivery of a devalued outcome contingent on an action can suppress the action independently of action–outcome encoding. Accumbens shell lesions, on the other hand, did not impair sensitivity to outcome devaluation on either the extinction test or the rewarded test, whereas accumbens core lesions abolished sensitivity to

devaluation on both tests (Corbit *et al.*, 2001). Sensitivity to contingency degradation, however, was not affected by either lesion, demonstrating that, after accumbens lesions, the rats were able to encode and to retrieve action–outcome representations.

The role of DA: mesolimbic vs. nigrostriatal

Ever since the pioneering studies on the phasic activity of DA neurons in monkeys, a common assumption in the field is that all DA cells behave in essentially the same way (Schultz, 1998a; Montague *et al.*, 2004). However, the available data, as well as the anatomical connectivity, suggest otherwise. In fact, the above analysis of functional heterogeneity in the striatum can be extended to the DA cells in the midbrain as well.

DA cells can be divided into two major groups: VTA and SNc. Although the projection from the VTA to accumbens has been the center of attention in the field of reward-related learning, the much more massive nigrostriatal pathway has been relatively neglected, with attention focused primarily on its role in Parkinson's disease. Current thinking on the role of DA in learning has been heavily influenced by the proposal that the phasic activity of DA cells reflects a reward prediction error (Ljungberg *et al.*, 1992; Schultz, 1998b). In the most common Pavlovian conditioning task used by Schultz and colleagues, these neurons fire in response to reward (US) but, with learning, the US-evoked activity is shifted to the CS. When the US is omitted after learning, the DA cells show a brief depression in activity at the expected time of its delivery (Waelti *et al.*, 2001; Fiorillo *et al.*, 2003; Tobler *et al.*, 2003). Such data form the basis of a variety of computational models (Schultz *et al.*, 1997; Schultz, 1998b; Brown *et al.*, 1999; Montague *et al.*, 2004).

Given multiple levels of control in the mechanisms of synthesis and release, the spiking of DA neurons cannot be equated with DA release, though one would expect these two measures to be highly correlated. Indeed, as shown by a recent study by Carelli and colleagues using fast-scan cyclic voltammetry, actual DA release in the accumbens core appears to be correlated with a prediction error in appetitive Pavlovian conditioning (Day *et al.*, 2007). They found a phasic DA signal in the accumbens core immediately after receipt of sucrose reward in Pavlovian autoshaping. After extended Pavlovian conditioning, however, this signal was no longer found after the reward itself, but shifted to the CS instead. This finding supports the original 'prediction error' hypothesis. It is also consistent with earlier work showing impaired performance of the Pavlovian CR after either DA receptor antagonism or DA depletion in the accumbens core (Di Ciano *et al.*, 2001; Parkinson *et al.*, 2002). However, one observation from the study is new and of considerable interest: after extended conditioning with a CS+ that predicts reward and a CS− that does not predict reward, a similar, though smaller, DA signal was also observed after the CS−, though it also showed a slight dip immediately (500–800 ms after cue onset) after the initial peak (Day *et al.*, 2007, fig. 4). By this stage in learning, animals almost never approach the CS−, but consistently approach the CS+. Thus, the phasic DA signal immediately after the predictor may not play a causal role in generating the approach response, as it is present even in the absence of the response. Whether such a signal is still necessary for learning the stimulus–reward contingency remains unclear, but the observed phasic response to the CS− is certainly not predicted by any of the current models.

Interestingly, local DA depletion does impair performance on this task (Parkinson *et al.*, 2002). Whereas a phasic DA signal is observed after the CS−, which does not generate CRs at all, abolishing both phasic and tonic DA by local depletion does impair the performance of

CRs. Such a pattern suggests that a phasic DA signal in the accumbens is not needed for performance of the Pavlovian CR, but may play a role in learning, while a slower, more tonic DA signal (presumably abolished in depletion studies) is more important for performance of the approach response (Cagniard *et al.*, 2006; Yin *et al.*, 2006b; Niv *et al.*, 2007). This possibility remains to be tested.

Although there is no direct evidence for a causal role of the phasic DA signal in learning, the 'prediction error' hypothesis has nevertheless attracted much attention, because it is precisely the type of teaching signal used in prominent models of learning, such as the Rescorla–Wagner model and its real-time extension the temporal difference reinforcement learning algorithm (Schultz, 1998b). According to this interpretation, appetitive learning is determined by the difference between received and expected reward (or between two temporally successive reward predictions). Such a teaching signal is regulated by negative feedback from all predictors of the reward (Schultz, 1998b). If no reward follows the predictor, then the negative feedback mechanism is unmasked as a dip in the activity of the DA neurons. Thus, learning involves the progressive reduction of the prediction error.

The elegance of the teaching signal in these models has perhaps distracted some from the anatomical reality. In the study by Day *et al.* (2007), the DA signal in the accumbens comes mostly from cells in the VTA, but it seems unlikely that other DA cells, with entirely different anatomical connectivity, would show the same response profile and provide the same signal. A gradient in what the DA cells signal is more likely, as DA cells project to different striatal regions with entirely different functions, and receive, in turn, distinct negative feedback signals from different striatal regions as well (Joel & Weiner, 2000; Wickens *et al.*, 2007). The mechanisms of uptake and degradation, as well as the presynaptic receptors that regulate release of DA, also show considerable variation across the striatum (Cragg *et al.*, 2002; Rice & Cragg, 2004, 2008; Wickens *et al.*, 2007).

We propose, therefore, that the mesoaccumbens pathway plays a more restricted role in Pavlovian learning, in acquiring the value of states and stimuli, whereas the nigrostriatal pathway is more important for instrumental learning, in acquiring the values of actions. That is, the phasic DA signal can encode different prediction errors, rather than a single prediction error, as is currently assumed. Three lines of evidence support this argument. First, genetic depletion of DA in the nigrostriatal pathway impairs the acquisition and performance of instrumental actions, whereas depletion of DA in the mesolimbic pathway does not (Sotak *et al.*, 2005; Robinson *et al.*, 2007). Second, DA cells in the SNc may encode the value of actions, similar to cells in their target striatal region (Morris *et al.*, 2006). Third, selective lesion of the nigrostriatal projection to the DLS impairs habit formation (Faure *et al.*, 2005).

Recent work by Palmiter and colleagues showed that genetically engineered DA-deficient mice are severely impaired in instrumental learning and performance, but their performance could be restored either by L-DOPA injection or by viral gene transfer to the nigrostriatal pathway (Sotak *et al.*, 2005; Robinson *et al.*, 2007). By contrast, DA restoration in the ventral striatum was not necessary to restore instrumental behavior. Although how DA signals enable instrumental learning remains an open question, one obvious possibility is that it could encode the value of self-initiated actions, i.e. how much reward is predicted given a particular course of action.

The dorsal striatum, as a whole, contains the highest expression of DA receptors in the brain, and receives the most massive dopaminergic projection. The DA projection to the DMS may play a different role in learning than the projection to the DLS, as these two regions differ significantly in the temporal profile of DA release, uptake and degradation (Wickens *et al.*, 2007). We hypothesize that the DA projection to the DMS from the medial SNc is critical for action–outcome learning, whereas the DA projection to the DLS from the lateral SNc is critical for habit formation. Should this be true, one should expect DA cells in the SNc to encode the error in reward prediction based on self-generated actions – instrumental prediction error – rather than that based on the CS. Preliminary evidence in support of this claim comes from a recent study by Morris *et al.*, who recorded from SNc neurons during an instrumental learning task (Morris *et al.*, 2006). Monkeys were trained to move their arms in response to a discriminative stimulus (S^D) that indicated the appropriate movement and the probability of reward. The S^D elicited phasic activity in the DA neurons corresponding to the action value based on the expected reward probability of a particular action. Most interestingly, although the DA response to the S^D increased with action value, the inverse was true of the DA response to the reward itself, consistent with the idea that these neurons were encoding a prediction error associated with that value. Not surprisingly, the primary striatal target of these cells, the caudate nucleus, is known to contain neurons that encode action values (Samejima *et al.*, 2005). It should be noted, however, that this study did not use behavioral tasks that unambiguously assess the value of actions. A clear prediction of our model is that phasic DA activity will accompany the performance of actions, even in the absence of an explicit S^D. For instance, we predict burst firing of nigral DA neurons at the time of a self-initiated action earning a reward.

In our view, whereas the mesoaccumbens DA signal reflects the value of the CS, the nigrostriatal signal, perhaps from those neurons projecting to the DMS, reflects the value of the action itself, or of any S^D that predicts this value. Moreover, both instrumental and Pavlovian learning appear to involve some form of negative feedback to control the effective teaching signal. In fact, the direct projections from the striatum to the midbrain DA neurons (Fig. 2) have long been proposed as the neural implementation of this type of negative feedback (Houk *et al.*, 1995), and the strength and nature of the inhibitory input may well vary considerably from region to region.

A prediction error, according to current models, is a teaching signal that determines how much learning occurs. So long as it is present, learning continues. However obvious this claim appears, a prediction error for action value, though syntactically similar to the Pavlovian prediction error, has unique features that have not been examined extensively. In traditional models like the Rescorla–Wagner model, which exclusively address Pavlovian conditioning (though with limited success), the key feature is the negative feedback that regulates prediction error. This output represents the acquired prediction, more specifically the 'sum' of all current predictors, as captured by the compound stimuli typically used in blocking experiments (Rescorla, 1988). It is this summing of available predictors to establish a global error term that is the chief innovation in this class of models. For instrumental actions, however, individual error terms seem more likely, for it is difficult to see how the negative feedback would present the value of multiple actions simultaneously when only one action can be performed at a time. Of course, a number of possible solutions do exist. For instance, given a particular state (experimentally implemented by a distinct S^D), the possible courses of actions could indeed be represented simultaneously as acquired predictions. But the chief difficulty with instrumental prediction errors has to do with the nature of the action itself. A Pavlovian prediction automatically follows the presentation of the stimulus, which is independent of the organism. An instrumental prediction error must address the element of control, because the prediction is itself action-contingent, and a deliberated action is emitted spontaneously based on

the animals' pursuit of the consequences of acting rather than elicited by antecedent stimuli. In the end, it is precisely a general neglect of the spontaneous nature of goal-directed actions, in both neuroscience and psychology, that has blurred the distinction between Pavlovian and instrumental learning processes, and the nature of the prediction errors involved. It remains to be established, therefore, what type of negative feedback signal, if any, regulates the acquisition of action values (Dayan & Balleine, 2002).

Finally, recent work has also implicated the nigrostriatal projection from the lateral SNc to DLS specifically in habit formation. Faure *et al.* selectively lesioned the DA cells projecting to DLS using 6-OHDA, and found that this manipulation has surprisingly little effect on the rate of lever pressing, though it impaired habit formation, as measured using outcome devaluation (Faure *et al.*, 2005). That is, lesioned animals responded in a goal-directed manner, even though, in a control group, the training generated habitual behavior insensitive to outcome devaluation. Local DA depletion, then, is similar to excitotoxic lesions of the DLS, in that both manipulations retard habit formation and favor the acquisition of goal-directed actions (Yin *et al.*, 2004). A phasic DA signal critical for habit formation is already well described by the effective reinforcement signal in contemporary temporal-difference reinforcement learning algorithms inspired by the work of Hull and Spence (Hull, 1943; Spence, 1947, 1960; Sutton & Barto, 1998).

Cortico-basal ganglia networks

So far we have discussed the functional heterogeneity within the striatum, yet it would be misleading to suggest that any striatal area could, say, translate the action–outcome contingency into the perfor-mance of an action all by itself. Rather the cerebral hemispheres are organized as iterating functional units consisting of cortico-basal ganglia networks (Swanson, 2000; Zahm, 2006). The striatum, being the entry station of the entire basal ganglia, serves as a unique hub in the cortico-basal ganglia network motif, capable of integrating cortical, thalamic and midbrain inputs. As described above, although it is a continuous structure, different striatal regions appear to participate in distinct functional networks, e.g. the accumbens acts as a hub in the limbic network and the DLS in the sensorimotor network. Due to the reentrant property of such networks, however, no one component of this structure is upstream or downstream in any absolute sense; e.g. the thalamocortical system is both the source of a major input to the striatum and the target of both the striato-pallidal and striato-nigral pathways.

Although parallel reentrant basal ganglia loops have long been recognized (Alexander *et al.*, 1986), we emphasize distinct functional roles of these circuits based on operationally defined representational structures and on interactions between circuits in generating integra-tive behaviors. On this basis, at least four such networks can be discerned: the limbic networks involving the shell and core of the accumbens, respectively; the associative network involving the associative striatum (DMS); and the sensorimotor network involving the sensorimotor striatum (DLS). Their functions range from medi-ating the control of appetitive Pavlovian URs and CRs to instrumental actions (Fig. 1).

As already mentioned, the ventral striatum consists mostly of the nucleus accumbens, which can be further divided into the shell and the core, each participating in a distinct functional network. The cortical (glutamatergic) projections to the shell arise from infralimbic, central and lateral orbital cortices, whereas the projections to the core arise from more dorsal midline regions of prefrontal cortex like the ventral

~DMS
A-O (instrumental)
goal-directed actions

~DLS
S-R (instrumental)
habits
skills
behavioral
sequencing

~Core
S-O (Pavlovian)
Preparatory CRs
Anticipatory approach

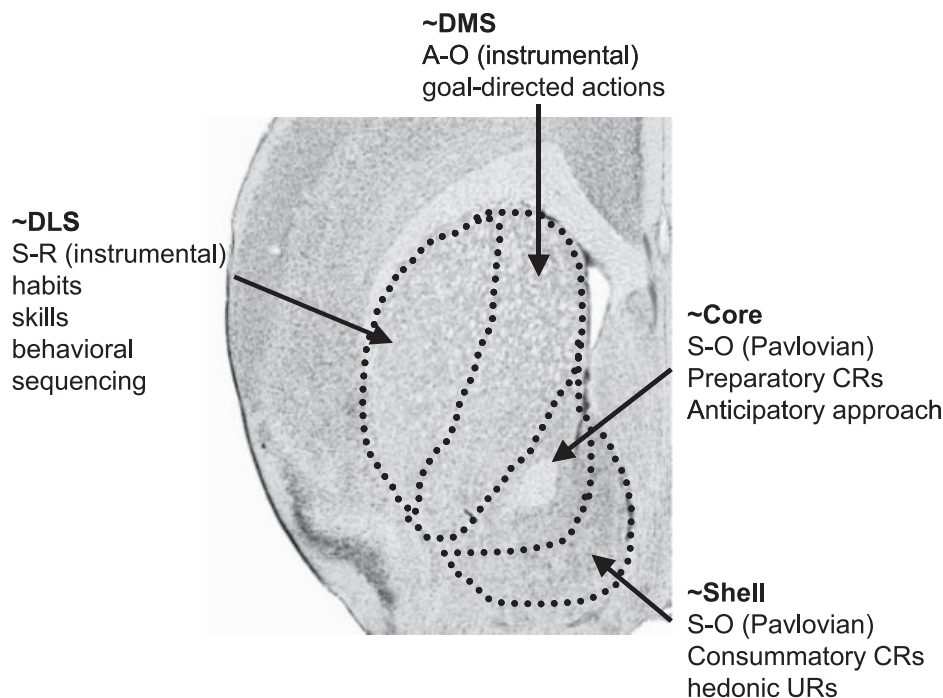**~Shell**
S-O (Pavlovian)
Consummatory CRs
hedonic URs

FIG. 1. Major functional domains of the striatum. An illustration of the striatum from a coronal section showing half of the brain (Paxinos & Franklin, 2003). Note that these four functional domains are anatomically continuous, and roughly correspond to what are commonly known as nucleus accumbens shell and core (limbic striatum), dorsomedial (DMS, association) striatum and dorsolateral (DLS, sensorimotor) striatum. We have not included other ventral striatal regions (e.g. areas posterior to the nucleus accumbens), which are not well understood. According to our framework, these limbic striatal areas should be broadly similar to the accumbens in function. A–O, action–outcome; CR, conditional response; S–O, stimulus–outcome; UR, unconditional response.

and dorsal prelimbic and anterior cingulate cortices (Groenewegen *et al.*, 1990; Zahm, 2000, 2006). Within these function networks evidence reviewed above suggests that the shell is involved in URs to rewards and the acquisition of consummatory CRs; the core in exploratory behavior, particularly the acquisition and expression of Pavlovian approach responses. At least two major networks, then, can be discerned within the larger ventral or limbic cortico-basal ganglia network, one for consummatory and the other for preparatory behaviors and their modification by Pavlovian conditioning (Fig. 1).

The dorsal striatum likewise can be divided into at least two major regions, associative and sensorimotor, with a distinct functional network associated with each. The associative striatum (caudate and parts of the anterior putamen in primates) contains neurons that fire in anticipation of response-contingent rewards and change their firing according to the magnitude of the expected reward (Hikosaka *et al.*, 1989; Hollerman *et al.*, 1998; Kawagoe *et al.*, 1998). In the associative network, the prefrontal and parietal association cortices and their target in the DMS are involved in transient memory, both prospective, in the form of outcome expectancies, and retrospective, as a record of recent efference copies (Konorski, 1967). The sensorimotor level, on the other hand, comprises the sensorimotor cortices and their targets in the basal ganglia. The outputs of this circuit are directed at motor cortices and brain stem motor networks. Neural activity in the sensorimotor striatum is generally not modulated by reward expectancy, displaying more movement-related activity than neurons in the associative striatum (Kanazawa *et al.*, 1993; Kimura *et al.*, 1993; Costa *et al.*, 2004). Finally, in addition to the medial-lateral gradient, there is significant functional heterogeneity along the anterior-posterior axis of the dorsal striatum, though sufficient data are not currently available to permit any detailed classification (Yin *et al.*, 2005b).

Studies have so far only focused on the cortical and striatal components of these networks. In general, lesions of a cortical area have similar effects as lesions of its striatal target (Balleine & Dickinson, 1998; Corbit & Balleine, 2003; Yin *et al.*, 2005b), but other components in the network could subserve similar functions. For example, lesions of the mediodorsal nucleus of the thalamus, a component of the associative network, were found to abolish sensitivity to outcome devaluation and contingency degradation in much the same way as lesions to the DMS and the prelimbic cortex (Corbit *et al.*, 2003). Thus, although our general model predicts similar behavioral deficits after damage to each component of a network, it also suggests, for any given structure like pallidum or thalamus, multiple functional domains.

Interaction between networks

Under most conditions, Pavlovian and instrumental learning appear to take place in parallel. Phenomena like PIT, however, demonstrate the extent to which these otherwise distinct processes can interact. Having delineated independent functional systems, the next step is to understand how these systems are coordinated to generate behavior. One attractive proposal, in accord with recent anatomical work, is that the networks outlined above are hierarchically organized, each serving as a labile, functional intermediary in the hierarchy, allowing information to propagate from one level to the next. In particular, the recently discovered spiraling connections between the striatum and the midbrain suggest an anatomical organization that can potentially implement interactions between networks (Fig. 2). As observed by Haber and colleagues, striatal neurons send direct inhibitory projections to DA neurons from
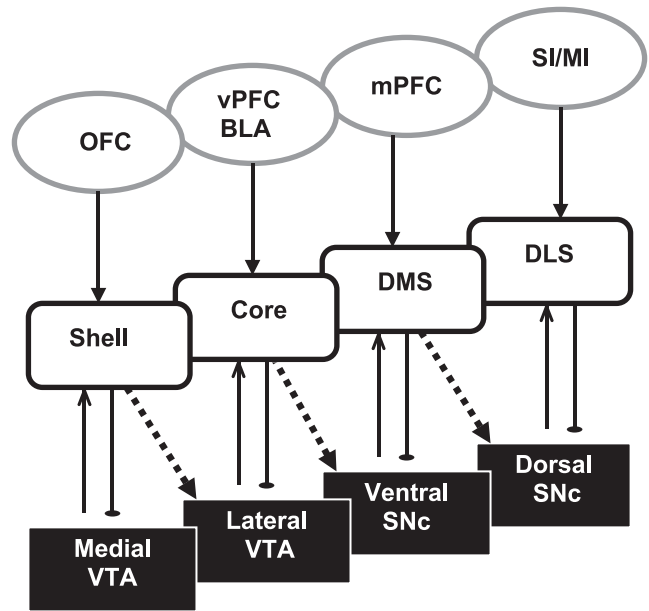

FIG. 2. The cortico-basal ganglia networks. An illustration of the major corticostriatal projections and dopaminergic projections in terms of the four major cortico-basal ganglia networks and their corresponding behavioral functions. Pallidal, thalamic and other structures have been omitted for the sake of clarity. Emphasis is placed on the spiraling midbrain–striatum–midbrain projections, which allows information to be propagated forward in a hierarchical manner. Note that this is only one possible neural implementation; interactions via different thalamo-cortico-thalamic projections are also possible (Haber, 2003). BLA, basolateral amygdale complex; core, nucleus accumbens core; DLS, dorsolateral striatum; DMS, dorsomedial striatum; mPFC, medial prefrontal cortex; OFC, orbitofrontal cortex; shell, nucleus accumbens shell; SI/MI, primary sensory and motor cortices; SNc, substantia nigra pars compacta; vPFC, ventral prefrontal cortex; VTA, ventral tegmental area.

which they receive reciprocal DA projections, and also project to DA neurons, which in turn project to a different striatal area (Haber *et al.*, 2000). These projections allow feedforward propagation of information in only one direction, from the limbic networks to associative and sensorimotor networks. For example, a Pavlovian prediction (acquired value of the CS) could reduce the effective teaching signal at the limbic level, while coincidentally potentiating the DA signal at the next level. The cancellation of the effective teaching signal is normally implemented by a negative feedback signal via an inhibitory projection, for example from the γ-aminobutyric acid (GABA)ergic medium spiny projection neurons from the striatum to the DA neurons. Meanwhile, as suggested by the anatomical organization (Haber *et al.*, 2000; Haber, 2003), the potentiation of the DA signal for the neighboring cortico-basal ganglia network (the next level in the hierarchy) could be implemented via disinhibitory projections (i.e. GABAergic striatal projection neurons to nigral GABAergic interneurons to DA neurons). Thus, the learned value of the limbic network can be transferred to the associative network, allowing behavioral adaptation to be refined and amplified with each iteration (Ashby, 1960). This model predicts, therefore, the progressive involvement of different neural networks during different stages of learning, a suggestion supported by a variety of data (Jueptner *et al.*, 1997b; Miyachi *et al.*, 1997, 2002; Yin, 2004; Everitt & Robbins, 2005; Yin & Knowlton, 2005; Belin & Everitt, 2008).

Phenomena that require the interaction of distinct functional processes, such as PIT, provide a fertile testing ground for models of

this kind. Indeed, the hierarchical model is in accord with recent experimental findings on PIT. According to the model, Pavlovian-instrumental interactions are mediated by reciprocal connections between the striatum and DA neurons. DA appears to be critical for general transfer, which is abolished by DA antagonists and local inactivation of the VTA (Dickinson *et al.*, 2000; Murschall & Hauber, 2006), whereas local infusion of amphetamine, which presumably increases DA levels, into the accumbens can significantly enhance it (Wyvell & Berridge, 2000). On the other hand, the role of ventral striatal DA in specific transfer is less clear. Some evidence suggests that it might be spared after inactivation of the VTA (Corbit *et al.*, 2007) but, as Corbit & Janak (2007) reported recently, specific transfer is abolished by inactivation of the DLS, suggesting that this aspect of stimulus control over action selection might involve the nigrostriatal projection (Corbit & Janak, 2007). Agreeing with the hierarchical perspective, Corbit & Janak (2007) also found that, whereas DLS inactivation abolished the selective excitatory effect of Pavlovian cues (much as has been observed after lesions of accumbens shell by Corbit *et al.*, 2001), inactivation of the DMS abolished only the outcome-selectivity of the transfer whilst appearing to preserve the general excitatory effect of these cues, a trend also observed after lesions of mediodorsal thalamus, which is part of the associative cortico-basal ganglia network (Ostlund & Balleine, 2008). Based on these preliminary results, the DMS appears to mediate only specific transfer, whereas the DLS could be necessary for both the specific and general excitatory effects of Pavlovian cues on instrumental actions.

Interestingly, the limbic striatum projects extensively to DA cells that project to the dorsal striatum (Nauta *et al.*, 1978; Nauta, 1989); the dopaminergic projections to the striatum and the striatal projections back to the midbrain are highly asymmetrical (Haber, 2003). The limbic striatum receives limited input from DA neurons yet sends extensive output to a much greater set of DA neurons, and the opposite is true of the sensorimotor striatum. Thus, the limbic networks are in a perfect position to control the associative and sensorimotor networks. Here the neuroanatomy agrees with behavioral data that the Pavlovian facilitation of instrumental behavior is much stronger than the reverse; indeed, considerable evidence suggests that instrumental actions tend to inhibit, rather than excite, Pavlovian CRs – a finding that still awaits a neurobiological explanation (Ellison & Konorski, 1964; Williams, 1965).

Conclusions

The hierarchical model discussed here, it should be noted, is very different from others that rely exclusively on the cortex and long-range connections between cortical areas (Fuster, 1995). It incorporates the known components and connectivity of the brain, rather than viewing it as a potpourri of cortical modules that, in some unspecified manner, implement a wide range of cognitive functions. It also avoids assumptions, inherited from 19th century neurology, that the cerebral cortex in general, and the prefrontal cortex in particular, somehow forms a 'higher' homuncular unit that controls the entire brain (Miller & Cohen, 2001).

Furthermore, several specific predictions can be derived from the present model. (i) There should be distinct prediction errors for self-generated actions and for states/stimuli with properties reflecting their different neural substrates and functional roles. (ii) The pallidal and thalamic components of each discrete cortico-basal ganglia network are also expected to be necessary for the type of behavioral control hypothesized for each network, not just the cortical and striatal

components. (iii) There should be a progressive involvement of different neural networks during different stages of learning. (iv) Accumbens activity can directly control DA neurons and, in turn, dorsal striatal activity. Based on a report by Holland (2004) suggesting that PIT increases with instrumental training, this 'limbic' control of the associative and sensorimotor networks is expected to strengthen with extended training.

Without detailed data, it is still too early to offer a formal account of the hierarchical model. Nevertheless, the above discussion should make it clear that current versions of the mesoaccumbens reward hypothesis rest on problematic assumptions about the nature of the reward process and the use of inadequate behavioral measures. Unifying principles, always the goal of the scientific enterprise, can only be founded on the reality of experimental data, however unwieldy these may be. Because the function of the brain is, ultimately, the generation and control of behavior, detailed behavioral analysis will be the key to understanding neural processes, much as a thorough description of innate and acquired immunity permits the elucidation of the immune system. Though seemingly a truism, it can hardly be overemphasized that we can understand brain mechanisms to the extent that their functions are described and measured with precision. When the study of neural function is based on experimentally established psychological capacities, for example the representation of action–outcome and stimulus–outcome contingencies, the known anatomical organization as well as physiological mechanisms are seen in a new light, leading to the formulations of new hypotheses and the design of new experiments. As an initial step in this direction, we hope that the framework discussed here will serve as a useful starting point for future investigation.

Acknowledgements

Abbreviations

CaMKII, calcium/calmodulin-dependent protein kinase II; CR, conditional response; CS, conditional stimulus; DA, dopamine; DLS, dorsolateral striatum; DMS, dorsomedial striatum; GABA, γ-aminobutyric acid; PIT, Pavlovian-instrumental transfer; S^D, discriminative stimulus; SNc, substantia nigra pars compacta; UR, unconditional response; US, unconditional stimulus; VTA, ventral tegmental area.

References

Adams, C.D. (1982) Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q. J. Exp. Psychol.*, **33b**, 109–122.

Adams, C.D. & Dickinson, A. (1981) Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol.*, **33**, 109–122.

Alexander, G.E., DeLong, M.R. & Strick, P.L. (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.*, **9**, 357–381.

Ashby, W.R. (1960) *Design for a Brain*, second Edition. Chapman & Hall, London.

Atallah, H.E., Lopez-Paniagua, D., Rudy, J.W. & O'Reilly, R.C. (2007) Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. *Nat. Neurosci.*, **10**, 126–131.

Balleine, B.W. (2001) Incentive processes in instrumental conditioning. In Mowrer, R.R. & Klein, S.B. (Eds), *Handbook of contemporary learning theories*. Lawrence Erlbaum Associates, Inc., Publishers, Mahwah, NJ, US, pp. 307–366.

Balleine, B.W. (2005) Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. *Physiol. Behav.*, **86**, 717–730.

Balleine, B.W. & Dickinson, A. (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, **37**, 407–419.

Balleine, B.W. & Killcross, S. (2006) Parallel incentive processing: an integrated view of amygdala function. *Trends Neurosci.*, **29**, 272–279.

Belin, D. & Everitt, B.J. (2008) Cocaine Seeking Habits Depend upon Dopamine-Dependent Serial Connectivity Linking the Ventral with the Dorsal Striatum. *Neuron*, **57**, 432–441.

Berke, J.D. & Hyman, S.E. (2000) Addiction, dopamine, and the molecular mechanisms of memory. *Neuron*, **25**, 515–532.

Berridge, K.C. & Robinson, T.E. (1998) What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res. Brain Res. Rev.*, **28**, 309–369.

Bolles, R. (1972) Reinforcement, expectancy, and learning. *Psychol. Rev.*, **79**, 394–409.

de Borchgrave, R., Rawlins, J.N., Dickinson, A. & Balleine, B.W. (2002) Effects of cytotoxic nucleus accumbens lesions on instrumental conditioning in rats. *Exp. Brain Res.*, **144**, 50–68.

Brown, P.L. & Jenkins, H.M. (1968) Auto-shaping the pigeon's key peck. *J. Exp. Anal. Behav.*, **11**, 1–8.

Brown, J., Bullock, D. & Grossberg, S. (1999) How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *J. Neurosci.*, **19**, 10502–10511.

Cagniard, B., Beeler, J.A., Britt, J.P., McGehee, D.S., Marinelli, M. & Zhuang, X. (2006) Dopamine scales performance in the absence of new learning. *Neuron*, **51**, 541–547.

Cardinal, R.N. & Cheung, T.H. (2005) Nucleus accumbens core lesions retard instrumental learning and performance with delayed reinforcement in the rat. *BMC Neurosci.*, **6**, 9.

Cardinal, R.N., Parkinson, J.A., Hall, J. & Everitt, B.J. (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.*, **26**, 321–352.

Cheer, J.F., Aragona, B.J., Heien, M.L., Seipel, A.T., Carelli, R.M. & Wightman, R.M. (2007) Coordinated accumbal dopamine release and neural activity drive goal-directed behavior. *Neuron*, **54**, 237–244.

Colwill, R.M. & Rescorla, R.A. (1986) Associative structures in instrumental learning. In Bower, G. (Ed.), *The Psychology of Learning and Motivation*. Academic Press, New York, pp. 55–104.

Corbit, L.H. & Balleine, B.W. (2003) The role of prelimbic cortex in instrumental conditioning. *Behav. Brain Res.*, **146**, 145–157.

Corbit, L.H. & Janak, P.H. (2007) Inactivation of the lateral but not medial dorsal striatum eliminates the excitatory impact of Pavlovian stimuli on instrumental responding. *J. Neurosci.*, **27**, 13977–13981.

Corbit, L.H., Muir, J.L. & Balleine, B.W. (2001) The role of the nucleus accumbens in instrumental conditioning: Evidence of a functional dissociation between accumbens core and shell. *J. Neurosci.*, **21**, 3251–3260.

Corbit, L.H., Muir, J.L. & Balleine, B.W. (2003) Lesions of mediodorsal thalamus and anterior thalamic nuclei produce dissociable effects on instrumental conditioning in rats. *Eur. J. Neurosci.*, **18**, 1286–1294.

Corbit, L.H., Janak, P.H. & Balleine, B.W. (2007) General and outcome-specific forms of Pavlovian-instrumental transfer: the effect of shifts in motivational state and inactivation of the ventral tegmental area. *Eur. J. Neurosci.*, **26**, 3141–3149.

Costa, R.M., Cohen, D. & Nicolelis, M.A. (2004) Differential corticostriatal plasticity during fast and slow motor skill learning in mice. *Curr. Biol.*, **14**, 1124–1134.

Cragg, S.J., Hille, C.J. & Greenfield, S.A. (2002) Functional domains in dorsal striatum of the nonhuman primate are defined by the dynamic behavior of dopamine. *J. Neurosci.*, **22**, 5705–5712.

Dalley, J.W., Laane, K., Theobald, D.E., Armstrong, H.C., Corlett, P.R., Chudasama, Y. & Robbins, T.W. (2005) Time-limited modulation of appetitive Pavlovian memory by D1 and NMDA receptors in the nucleus accumbens. *Proc Natl Acad Sci U S A*, **102**, 6189–6194.

Davis, J. & Bitterman, M.E. (1971) Differential reinforcement of other behavior (DRO): A yoked-control comparison. *J. Exp. Anal. Behav.*, **15**, 237–241.

Day, J.J. & Carelli, R.M. (2007) The nucleus accumbens and Pavlovian reward learning. *Neuroscientist*, **13**, 148–159.

Day, J.J., Wheeler, R.A., Roitman, M.F. & Carelli, R.M. (2006) Nucleus accumbens neurons encode Pavlovian approach behaviors: evidence from an autoshaping paradigm. *Eur. J. Neurosci.*, **23**, 1341–1351.

Day, J.J., Roitman, M.F., Wightman, R.M. & Carelli, R.M. (2007) Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.*, **10**, 1020–1028.

Dayan, P. & Balleine, B.W. (2002) Reward, motivation, and reinforcement learning. *Neuron*, **36**, 285–298.

Delgado, M.R., Stenger, V.A. & Fiez, J.A. (2004) Motivation-dependent responses in the human caudate nucleus. *Cereb. Cortex*, **14**, 1022–1030.

Delgado, M.R., Miller, M.M., Inati, S. & Phelps, E.A. (2005) An fMRI study of reward-related probability learning. *Neuroimage*, **24**, 862–873.

Di Ciano, P., Cardinal, R.N., Cowell, R.A., Little, S.J. & Everitt, B.J. (2001) Differential involvement of NMDA, AMPA/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of pavlovian approach behavior. *J. Neurosci.*, **21**, 9471–9477.

Dickinson, A. (1985) Actions and habits: the development of behavioural autonomy. *Philos. Trans. R. Soc.*, **308**, 67–78.

Dickinson, A. (1994) Instrumental Conditioning. In Mackintosh, N.J. (Ed.), *Animal Learning and Cognition*. Academic, Orlando, pp. 45–79.

Dickinson, A. & Balleine, B. (1993) Actions and responses: The dual psychology of behaviour. In Eilan, N., McCarthy, R.A. & Brewer, B. (Eds), *Spatial Representation: Problems in Philosophy and Psychology*. Blackwell Publishers Inc, Malden, MA, US, pp. 277–293.

Dickinson, A. & Balleine, B. (2002) The role of learning in the operation of motivational systems. In Pashler, H. & Gallistel, R., (Eds), *Steven's Handbook of Experimental Psychology (3rd ed.), Vol. 3: Learning, Motivation, and Emotion*. John Wiley & Sons, Inc, New York, NY, pp. 497–533.

Dickinson, A. & Charnock, D.J. (1985) Contingency effects with maintained instrumental reinforcement. *Q. J. Exp. Psychol. B*, **37**, 397–416.

Dickinson, A. & Dearing, M.F. (1979) Appetitive-aversive interactions and inhibitory processes. In Dickinson, A. & Boakes, R.A. (Eds), *Mechanism of Learning and Motivation*. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 203–231.

Dickinson, A., Campos, J., Varga, Z.I. & Balleine, B. (1996) Bidirectional instrumental conditioning. *Q. J. Exp. Psychol B*, **49**, 289–306.

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

Ellison, G.D. & Konorski, J. (1964) Separation of the salivary and motor responses in instrumental conditioning. *Science*, **146**, 1071–1072.

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

Faure, A., Haberland, U., Conde, F. & El Massioui, N. (2005) Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *J. Neurosci.*, **25**, 2771–2780.

Fiorillo, C.D., Tobler, P.N. & Schultz, W. (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, **299**, 1898–1902.

Fuster, J.M. (1995). *Memory in the Cerebral Cortex*. MIT press, Cambridge.

Gallistel, C.R., Fairhurst, S. & Balsam, P. (2004) The learning curve: implications of a quantitative analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 13124–13131.

Goto, Y. & Grace, A.A. (2005) Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nat. Neurosci.*, **8**, 805–812.

Grace, A.A., Floresco, S.B., Goto, Y. & Lodge, D.J. (2007) Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. *Trends Neurosci.*, **30**, 220–227.

Groenewegen, H.J., Berendse, H.W., Wolters, J.G. & Lohman, A.H. (1990) The anatomical relationship of the prefrontal cortex with the striatopallidal system, the thalamus and the amygdala: evidence for a parallel organization. *Prog. Brain Res.*, **85**, 95–116. discussion 116-118.

Guthrie, E.R. (1935) *The Psychology of Learning*. Harpers, New York.

Haber, S.N. (2003) The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat.*, **26**, 317–330.

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

Hammond, L.J. (1980) The effect of contingency upon the appetitive conditioning of free-operant behavior. *J. Exp. Anal. Behav.*, **34**, 297–304.

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

Haruno, M. & Kawato, M. (2006b) Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *J. Neurophysiol.*, **95**, 948–959.

Haruno, M., Kuroda, T., Doya, K., Toyama, K., Kimura, M., Samejima, K., Imamizu, H. & Kawato, M. (2004) A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. *J. Neurosci.*, **24**, 1660–1665.

Hernandez, P.J., Sadeghian, K. & Kelley, A.E. (2002) Early consolidation of instrumental learning requires protein synthesis in the nucleus accumbens. *Nat. Neurosci.*, **5**, 1327–1331.

Hernandez, P.J., Andrzejewski, M.E., Sadeghian, K., Panksepp, J.B. & Kelley, A.E. (2005) AMPA/kainate, NMDA, and dopamine D1 receptor function in the nucleus accumbens core: a context-limited role in the encoding and consolidation of instrumental memory. *Learn. Mem.*, **12**, 285–295.

Hershberger, W.A. (1986) An approach through the looking glass. *Anim. Learn. Behav.*, **14**, 443–451.

Heyes, C.M. & Dawson, G.R. (1990) A demonstration of observational learning in rats using a bidirectional control. *Q. J. Exp. Psychol.*, **42**, 59–71.

Hikosaka, O., Sakamoto, M. & Usui, S. (1989) Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J. Neurophysiol.*, **61**, 814–832.

Holland, P.C. (1979) Differential effects of omission contingencies on various components of Pavlovian appetitive conditioned responding in rats. *J. Exp. Psychol. Anim. Behav. Process.*, **5**, 178–193.

Holland, P.C. (2004) Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *J. Exp. Psychol. Anim. Behav. Process.*, **30**, 104–117.

Holland, P.C. & Rescorla, R.A. (1975) The effect of two ways of devaluing the unconditioned stimulus after first- and second-order appetitive conditioning. *J. Exp. Psychol. Anim. Behav. Process.*, **1**, 355–363.

Hollerman, J.R., Tremblay, L. & Schultz, W. (1998) Influence of reward expectation on behavior-related neuronal activity in primate striatum. *J. Neurophysiol.*, **80**, 947–963.

Houk, J.C., Adams, J.L. & Barto, A.G. (1995) A model of how the basal ganglia generates and uses neural signals that predict reinforcement. In Houk, J.C., Davis, J.L. & Beiser, D.G. (Eds), *Models of Information Processing in the Basal Ganglia*. MIT Press, Cambridge, MA, pp. 249–270.

Hull, C. (1943). *Principles of Behavior*. Appleton-Century-Crofts, New York.

Hyman, S.E., Malenka, R.C. & Nestler, E.J. (2006) Neural mechanisms of addiction: the role of reward-related learning and memory. *Annu. Rev. Neurosci.*, **29**, 565–598.

Jedynak, J.P., Uslaner, J.M., Esteban, J.A. & Robinson, T.E. (2007) Methamphetamine-induced structural plasticity in the dorsal striatum. *Eur. J. Neurosci.*, **25**, 847–853.

Joel, D. & Weiner, I. (1994) The organization of the basal ganglia-thalamocortical circuits: open interconnected rather than closed segregated. *Neuroscience*, **63**, 363–379.

Joel, D. & 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**, 451–474.

Jueptner, M., Frith, C.D., Brooks, D.J., Frackowiak, R.S. & Passingham, R.E. (1997a) Anatomy of motor learning. II. Subcortical structures and learning by trial and error. *J. Neurophysiol.*, **77**, 1325–1337.

Jueptner, M., Stephan, K.M., Frith, C.D., Brooks, D.J., Frackowiak, R.S. & Passingham, R.E. (1997b) Anatomy of motor learning. I. Frontal cortex and attention to action. *J. Neurophysiol.*, **77**, 1313–1324.

Kanazawa, I., Murata, M. & Kimura, M. (1993) Roles of dopamine and its receptors in generation of choreic movements. *Adv. Neurol.*, **60**, 107–112.

Kawagoe, R., Takikawa, Y. & Hikosaka, O. (1998) Expectation of reward modulates cognitive signals in the basal ganglia. *Nat. Neurosci.*, **1**, 411–416.

Kimura, M., Aosaki, T. & Ishida, A. (1993) Neurophysiological aspects of the differential roles of the putamen and caudate nucleus in voluntary movement. *Adv. Neurol.*, **60**, 62–70.

Konorski, J. (1967) *Integrative Activity of the Brain*. University of Chicago Press, Chicago.

Lerchner, A., La Camera, G. & Richmond, B. (2007) Knowing without doing. *Nat. Neurosci.*, **10**, 15–17.

Ljungberg, T., Apicella, P. & Schultz, W. (1992) Responses of monkey dopamine neurons during learning of behavioral reactions. *J. Neurophysiol.*, **67**, 145–163.

Lohrenz, T., McCabe, K., Camerer, C.F. & Montague, P.R. (2007) Neural signature of fictive learning signals in a sequential investment task. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 9493–9498.

Lovibond, P.F. (1983) Facilitation of instrumental behavior by a Pavlovian appetitive conditioned stimulus. *J. Exp. Psychol. Anim. Behav. Process.*, **9**, 225–247.

Mackintosh, N.J. (1974) *The Psychology of Animal Learning*. Academic Press, London.

Miller, R. (1981) *Meaning and Purpose in the Intact Brain*. Oxford University Press, New York.

Miller, E.K. & Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.*, **24**, 167–202.

Miyachi, S., Hikosaka, O., Miyashita, K., Karadi, Z. & Rand, M.K. (1997) Differential roles of monkey striatum in learning of sequential hand movement. *Exp. Brain Res.*, **115**, 1–5.

Miyachi, S., Hikosaka, O. & Lu, X. (2002) Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Exp. Brain Res.*, **146**, 122–126.

Montague, P.R., Hyman, S.E. & Cohen, J.D. (2004) Computational roles for dopamine in behavioural control. *Nature*, **431**, 760–767.

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

Murschall, A. & Hauber, W. (2006) Inactivation of the ventral tegmental area abolished the general excitatory influence of Pavlovian cues on instrumental performance. *Learn. Mem.*, **13**, 123–126.

Nauta, W.J.H. (1989) Reciprocal links of the corpus striatum with the cerebral cortex and limbic system: A common substrate for movement and thought? In Mueller, J. (Ed), *Neurology and Psychiatry: A Meeting of Minds*. Karger, Basel, pp. 43–63.

Nauta, W.J., Smith, G.P., Faull, R.L. & Domesick, V.B. (1978) Efferent connections and nigral afferents of the nucleus accumbens septi in the rat. *Neuroscience*, **3**, 385–401.

Niv, Y., Daw, N.D., Joel, D. & Dayan, P. (2007) Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl)*, **191**, 507–520.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K. & Dolan, R.J. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, **304**, 452–454.

Ostlund, S.B. & Balleine, B.W. (2008) Differential involvement of the basolateral amygdala and mediodorsal thalamus in instrumental action selection. *J. Neurosci.*, **28**, 4398–4405.

Parkinson, J.A., Olmstead, M.C., Burns, L.H., Robbins, T. W. & 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**, 2401–2411.

Parkinson, J.A., Willoughby, P.J., Robbins, T.W. & Everitt, B.J. (2000) Disconnection of the anterior cingulate cortex and nucleus accumbens core impairs Pavlovian approach behavior: further evidence for limbic cortical-ventral striatopallidal systems. *Behav. Neurosci.*, **114**, 42–63.

Parkinson, J.A., Dalley, J.W., Cardinal, R.N., Bamford, A., Fehnert, B., Lachenal, G., Rudarakanchana, N., Halkerston, K.M., Robbins, T.W. & Everitt, B.J. (2002) Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. *Behav. Brain Res.*, **137**, 149–163.

Paxinos, G. & Franklin, K. (2003) *The Mouse Brain in Stereotaxic Coordinates*. Academic Press, New York.

Pecina, S., Smith, K.S. & Berridge, K.C. (2006) Hedonic hot spots in the brain. *Neuroscientist*, **12**, 500–511.

Pithers, R.T. (1985) The roles of event contingencies and reinforcement in human autoshaping and omission responding. *Learning and Motivation.*, **16**, 210–237.

Pothuizen, H.H., Jongen-Relo, A.L., Feldon, J. & Yee, B.K. (2005) Double dissociation of the effects of selective nucleus accumbens core and shell lesions on impulsive-choice behaviour and salience learning in rats. *Eur. J. Neurosci.*, **22**, 2605–2616.

Rescorla, R.A. (1968) Probability of shock in the presence and absence of CS in fear conditioning. *J. Comp. Physiol. Psychol.*, **66**, 1–5.

Rescorla, R.A. (1988) Behavioral studies of Pavlovian conditioning. *Annu. Rev. Neurosci.*, **11**, 329–352.

Rescorla, R.A. & Solomon, R.L. (1967) Two-process learning theory: relationships between Pavlovian conditioning and instrumental learning. *Psychol. Rev.*, **74**, 151–182.

Restle, F. (1957) Discrimination of cues in mazes: a resolution of the "place-vs.-response" question. *Psychol. Rev.*, **64**, 217.

Reynolds, J.N. & Wickens, J.R. (2002) Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw.*, **15**, 507–521.

Rice, M.E. & Cragg, S.J. (2004) Nicotine amplifies reward-related dopamine signals in striatum. *Nat. Neurosci.*, **7**, 583–584.

Rice, M.E. & Cragg, S.J. (2008) Dopamine spillover after quantal release: Rethinking dopamine transmission in the nigrostriatal pathway. *Brain Res. Rev.*, in press. [doi: 10.1016/j.brainresrer.2008.02.004.]

Robinson, S., Rainwater, A.J., Hnasko, T.S. & Palmiter, R.D. (2007) Viral restoration of dopamine signaling to the dorsal striatum restores instrumental conditioning to dopamine-deficient mice. *Psychopharmacology (Berl)*, **191**, 567–578.

Roitman, M.F., Wheeler, R.A. & Carelli, R.M. (2005) Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. *Neuron*, **45**, 587–597.

Samejima, K., Ueda, Y., Doya, K. & Kimura, M. (2005) Representation of action-specific reward values in the striatum. *Science*, **310**, 1337–1340.

Schultz, W. (1998a) The phasic reward signal of primate dopamine neurons. *Adv. Pharmacol.*, **42**, 686–690.

Schultz, W. (1998b) Predictive reward signal of dopamine neurons. *J. Neurophysiol.*, **80**, 1–27.

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

Schwartz, B. & Gamzu, E. (1977) Pavlovian control of operant behavior. In Honig, W. & Staddon, J.E.R. (Eds), *Handbook of Operant Behavior*. Prentice Hall, New Jersey, pp. 53–97.

Sheffield, F.D. (1965) Relation between classical and instrumental conditioning. In Prokasy, W.F. (Ed), *Classical Conditioning*. Appleton-Century-Crofts, New York, pp. 302–322.

Skinner, B. (1938) *The Behavior of Organisms*. Appleton-Century-Crofts, New York.

Smith-Roe, S.L. & Kelley, A.E. (2000) Coincident activation of NMDA and dopamine D1 receptors within the nucleus accumbens core is required for appetitive instrumental learning. *J. Neurosci.*, **20**, 7737–7742.

Sotak, B.N., Hnasko, T.S., Robinson, S., Kremer, E.J. & Palmiter, R.D. (2005) Dysregulation of dopamine signaling in the dorsal striatum inhibits feeding. *Brain Res.*, **1061**, 88–96.

Spence, K. (1947) The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, **54**, 1–8.

Spence, K. (1960) *Behavior Theory and Learning*. Prentice-Hall, Englewood Cliffs, NJ.

Sutton, R.S. & Barto, A.G. (1998) *Reinforcement Learning*. MIT Press, Cambridge.

Swanson, L.W. (2000) Cerebral hemisphere regulation of motivated behavior. *Brain Res.*, **886**, 113–164.

Taha, S.A. & Fields, H.L. (2005) Encoding of palatability and appetitive behaviors by distinct neuronal populations in the nucleus accumbens. *J. Neurosci.*, **25**, 1193–1202.

Taha, S.A. & Fields, H.L. (2006) Inhibitions of nucleus accumbens neurons encode a gating signal for reward-directed behavior. *J. Neurosci.*, **26**, 217–222.

Thorndike, E.L. (1911) *Animal Intelligence: Experimental Studies*. Macmillan, New York.

Tobler, P.N., Dickinson, A. & Schultz, W. (2003) Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *J. Neurosci.*, **23**, 10402–10410.

Tobler, P.N., O'Doherty, J.P., Dolan, R.J. & Schultz, W. (2006) Human neural learning depends on reward prediction errors in the blocking paradigm. *J. Neurophysiol.*, **95**, 301–310.

Tobler, P.N., O'Doherty, J.P., Dolan, R.J. & Schultz, W. (2007) Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J. Neurophysiol.*, **97**, 1621–1632.

Trapold, M.A. & Overmier, J.B. (1972) The second learning process in instrumental learning. In Black, A.A. & Prokasy, W.F. (Eds), *Classical Conditioning II: Current Research and Theory*. Appleton-Century-Crofts, New York, pp. 427–452.

Tricomi, E.M., Delgado, M.R. & Fiez, J.A. (2004) Modulation of caudate activity by action contingency. *Neuron*, **41**, 281–292.

Waelti, P., Dickinson, A. & Schultz, W. (2001) Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, **412**, 43–48.

White, N.M. (1989) A functional hypothesis concerning the striatal matrix and patches: mediation of S-R memory and reward. *Life Sci.*, **45**, 1943–1957.

Wickens, J.R., Budd, C.S., Hyland, B.I. & Arbuthnott, G.W. (2007) Striatal contributions to reward and decision making: making sense of regional variations in a reiterated processing matrix. *Ann. N. Y. Acad. Sci.*, **1104**, 192–212.

Williams, D.R. (1965) Classical conditioning and incentive motivation. In Prokasy, W.F. (Ed), *Classical Conditioning*. Appleton-Century-Crofts, New York, pp. 340–357.

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

Wiltgen, B.J., Law, M., Ostlund, S., Mayford, M. & Balleine, B.W. (2007) The influence of Pavlovian cues on instrumental performance is mediated by CaMKII activity in the striatum. *Eur. J. Neurosci.*, **25**, 2491–2497.

Wyvell, C.L. & Berridge, K.C. (2000) Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: enhancement of reward "wanting" without enhanced "liking" or response reinforcement. *J. Neurosci.*, **20**, 8122–8130.

Yin, H.H. (2004) The role of the dorsal striatum in goal-directed actions. Doctoral dissertation, Department of Psychology. UCLA, Los Angeles.

Yin, H.H. & Knowlton, B.J. (2002) Reinforcer devaluation abolishes conditioned cue preference: evidence for stimulus-stimulus associations. *Behav. Neurosci.*, **116**, 174–177.

Yin, H.H. & Knowlton, B.J. (2004) Contributions of striatal subregions to place and response learning. *Learn. Mem.*, **11**, 459–463.

Yin, H.H. & Knowlton, B.J. (2005) Addiction and learning. In Wiers, R.W. & Stacy, A. (Eds), *Handbook of Implicit Cognition and Addiction*. Sage, Thousand Oaks, pp. 167–183.

Yin, H.H., Knowlton, B.J. & Balleine, B.W. (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.*, **19**, 181–189.

Yin, H.H., Knowlton, B.J. & Balleine, B.W. (2005a) Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *Eur. J. Neurosci.*, **22**, 505–512.

Yin, H.H., Ostlund, S.B., Knowlton, B.J. & Balleine, B.W. (2005b) The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.*, **22**, 513–523.

Yin, H.H., Knowlton, B.J. & Balleine, B.W. (2006a) Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav. Brain Res.*, **166**, 189–196.

Yin, H.H., Zhuang, X. & Balleine, B.W. (2006b) Instrumental learning in hyperdopaminergic mice. *Neurobiol. Learn. Mem.*, **85**, 283–288.

Zahm, D.S. (2000) An integrative neuroanatomical perspective on some subcortical substrates of adaptive responding with emphasis on the nucleus accumbens. *Neurosci. Biobehav. Rev.*, **24**, 85–105.

Zahm, D.S. (2006) The evolving theory of basal forebrain functional-anatomical 'macrosystems'. *Neurosci. Biobehav. Rev.*, **30**, 148–172.