Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action

Bernard W Balleine*[,1,2] and John P O'Doherty[3,4]

[1]Brain and Mind Research Institute, University of Sydney, Camperdown, NSW, Australia; [2]Departments of Psychology and Psychiatry, UCLA, Los Angeles, CA, USA; [3]Institute of Neuroscience and School of Psychology, Trinity College Dublin, Dublin, Ireland; [4]HSS and Computation and Neural Systems, California Institute of Technology, Pasadena, CA, USA

Recent behavioral studies in both humans and rodents have found evidence that performance in decision-making tasks depends on two different learning processes; one encoding the relationship between actions and their consequences and a second involving the formation of stimulus–response associations. These learning processes are thought to govern goal-directed and habitual actions, respectively, and have been found to depend on homologous corticostriatal networks in these species. Thus, recent research using comparable behavioral tasks in both humans and rats has implicated homologous regions of cortex (medial prefrontal cortex/medial orbital cortex in humans and prelimbic cortex in rats) and of dorsal striatum (anterior caudate in humans and dorsomedial striatum in rats) in goal-directed action and in the control of habitual actions (posterior lateral putamen in humans and dorsolateral striatum in rats). These learning processes have been argued to be antagonistic or competing because their control over performance appears to be all or none. Nevertheless, evidence has started to accumulate suggesting that they may at times compete and at others cooperate in the selection and subsequent evaluation of actions necessary for normal choice performance. It appears likely that cooperation or competition between these sources of action control depends not only on local interactions in dorsal striatum but also on the cortico-basal ganglia network within which the striatum is embedded and that mediates the integration of learning with basic motivational and emotional processes. The neural basis of the integration of learning and motivation in choice and decision-making is still controversial and we review some recent hypotheses relating to this issue.

INTRODUCTION

Although much has been discovered about the anatomy and pharmacology of the basal ganglia, there has been much less research systematically investigating their functions in adaptive behavior. The considerable interest in neuropathology, particularly of the dorsal striatum and its midbrain dopaminergic afferents, associated with disorders, such as obsessive compulsive disorder (Saint-Cyr *et al*, 1995), Parkinson's (Antonini *et al*, 2001) and Huntington's diseases (Joel, 2001), Tourette's syndrome (Marsh *et al*, 2004), and so on, has implicated this structural network in various aspects of action control. However, historically, these functions have been argued to be limited to the bottom-up regulation of motor movement through output to the pallidum, thalamus, and to motor and premotor cortices (Mink, 1996). More recently, however, interest has turned to potential role of the basal ganglia in the top-down or executive control of motor movement (Miller, 2008). This interest has largely been fueled by new, more detailed, models of the neuroanatomy that have linked feedforward inputs through the corticostriatal circuit with these feedback functions through a network of partially closed cortico-basal ganglia loops (Alexander and Crutcher, 1990b; Nambu, 2008).

In addition, this recent approach has led to integration of some previously poorly characterized aspects of basal ganglia function with motor control, particularly its role in decision-making based on affective or reward processes.

*Correspondence: Professor B Balleine, Brain and Mind Research Institute, University of Sydney, 100 Mallett Street, Camperdown, NSW 2050, Australia, Tel: +61 2 9114-4011, Fax: +61 2 9114-4034, E-mail: balleine@med.usyd.edu.au

Considerable research has long suggested that the ventral striatum has a critical role in the motivational control of action, a suggestion captured in the characterization of this region as a limbic–motor interface (Kelley, 2004; Mogenson *et al*, 1980). However, recent research has also implicated the dorsal striatum in the control of decision-making regulated by reward, particularly the role of the caudate or dorsomedial region of the striatum in the integration of reward-related processes with action control (Balleine *et al*, 2007c). This research suggests that the striatum has a much broader role in the control of executive functions than previously suspected and, indeed, appears to be centrally involved in functions long argued to depend solely on the regions of prefrontal cortex (Balleine *et al*, 2007a; Lauwereyns *et al*, 2002; Tanaka *et al*, 2006).

Reconciling the motor with the non-motor, cognitive aspects of basal ganglia function in the production and control of adaptive behavior has become a matter of considerable speculation and debate, and a number of theories have been advanced based around the new and developing description of the neuroanatomy and neuropharmacology of basal ganglia and its composite subregions (Bar-Gad *et al*, 2003; Hazy *et al*, 2007; Joel *et al*, 2002; McHaffie *et al*, 2005). However, it must be conceded that, although promissory, much of this theorizing has been driven by these structural advances with relatively little consideration of evidence (or lack of evidence) regarding the functions of the circuitry involved (Nambu, 2008).

In this review, we describe recent research from our laboratories that addresses this issue focusing particularly on collaborative projects through which we have started to assess homologous functions of corticostriatal circuitry in both executive and motor learning processes and in the motivational control of action in rodents and humans. We begin by considering behavioral evidence for multiple sources of action control and recent evidence implicating regions of dorsal striatum in the cognitive and sensorimotor control of actions. We then consider the evidence for two, apparently independent, sources of motivational control mediated by reward and by stimuli that predict reward and the role of distinct cortico-ventral striatal networks in these functions. Finally, we consider the role of basal ganglia generally in the integration of the learning and motivational processes through which courses of action are acquired, selected, and implemented to determine adaptive decision-making and choice.

MULTIPLE SOURCES OF ACTION CONTROL IN RODENTS AND HUMANS: GOALS AND HABITS

There is now considerable evidence to suggest that the performance of reward-related actions in both rats and humans reflects the interaction of two quite different learning processes, one controlling the acquisition of goal-directed actions, and the other the acquisition of habits. This evidence suggests that, in the goal-directed case, action selection is governed by an association between the response 'representation' and the 'representation' of the outcome engendered by those actions, whereas in the case of habit learning, action selection is controlled through learned stimulus–response (S–R) associations without any associative link to the outcome of those actions. As such, actions under goal-directed control are performed with regard to their consequences, whereas those under habitual control are more reflexive in nature, by virtue of their control by antecedent stimuli rather than their consequences. It should be clear, therefore, that goal-directed and habitual actions differ in two primary ways: (1) they differ in their sensitivity to changes in the value of the consequences previously associated with the action; and (2) they differ in their sensitivity to changes in the causal relationship between the action and those consequences. Generally, two kinds of experimental test have been used to establish these differences, referred to as outcome devaluation and contingency degradation, respectively.

Outcome Devaluation

It is now nearly 30 years since it was first demonstrated that rats are capable of encoding the consequences of their actions. In an investigation of the learning processes controlling instrumental conditioning in rats, Adams and Dickinson (1981) were able to demonstrate that, after they were trained to press a lever for sucrose, devaluing the sucrose by pairing its consumption with illness (induced by an injection of lithium chloride) caused a subsequent reduction in performance when the rats were again allowed to lever press in an extinction test; ie, in the absence of any feedback from the delivery of the now devalued outcome. In the absence of this feedback during the test, the reduction in performance suggested that the rats encoded the lever press–sucrose association during training and were able to integrate that learning with the experienced change in the value of the sucrose to alter their subsequent performance.

This demonstration was of central importance because, at the time, the available evidence suggested that lever-press acquisition was controlled solely by sensorimotor learning, involving a process of S–R association. Adams and Dickinson's (1981) finding was the first direct evidence to contradict this view and to suggest that animals are capable of a more elaborate form of encoding based on the response–outcome (R–O) association. It is important to recognize that evidence of R–O encoding did not necessarily imply that animals could not or did not learn by S–R association and, indeed, subsequent evidence has suggested that both processes can be encouraged depending on the training conditions. Adams (1981) found early on that the influence of devaluation on lever pressing was dependent on the amount of training; that, after a period of overtraining,

the lever pressing by rats appeared to be no longer sensitive to devaluation. This was consistent with the view that R–O learning dominated performance early after acquisition but gave way to an S–R process, as performance became more routine or habitual (see also Dickinson, 1994; Dickinson *et al*, 1995).

Importantly, very similar effects have been found in human subjects. In a recent study, for example, we (Tricomi *et al*, 2009) trained human subjects to press different buttons to gain access to symbols corresponding to small quantities of two different snack foods (one of which they were given to eat at the end of the session). When allowed to eat a particular snack food until satiated, thereby selectively devaluing that snack food, undertrained subjects subsequently reduced their performance of the action associated with the devalued snack food compared with that of an action associated with a non-devalued snack food in an extinction test. In contrast, after overtraining, performance was no longer sensitive to snack food devaluation and subjects responded similarly on both the action associated with the devalued outcome and the action associated with the non-devalued outcome.

Devaluation effects have also been demonstrated when assessed using choice between different actions. Rats trained on two different action–outcome associations have been found to alter their choice between actions in an extinction test after one or other outcome has been devalued either by taste aversion learning (Colwill and Rescorla, 1985) or by specific satiety induced by a period of free consumption of one or other training outcome (Balleine and Dickinson, 1998a, b). Valentin *et al* (2007) reported a similar devaluation effect on choice performance in humans, using a training procedure in which subjects were allowed to make stochastic choices between two icons, one paired with a high probability chocolate milk or low probability orange juice and, in a second pair of icons, one paired with high probability tomato juice and the other the low probability orange juice. When subsequently sated on the chocolate milk or tomato juice, their choice of the specific icon associated with the now devalued outcome was reduced, whereas choice of the icon paired with the other non-devalued outcome was not.

Thus, whether assessed using an undertrained free operant action or by means of choice between two actions, both rodents and humans show sensitivity to outcome devaluation. Furthermore, in both species overtraining produces insensitivity to outcome devaluation suggesting that performance has become habitual. These findings suggest, therefore, that both rodents and humans are capable of goal-directed and habitual forms of behavioral control.

Contingency Degradation

In addition to differences in associative structure, demonstrated by differential sensitivity to devaluation, the R–O and S–R learning processes appear also to be driven by different learning rules. Traditionally S–R learning was argued

to reflect the operation of a contiguity learning rule, subsequently referred to as Hebbian-learning. As one might expect, this proposed that S–R learning is governed by contiguous activation of sensory and motor processes (Hebb, 1949). On a pure contiguity account (Guthrie, 1935) that is all that is required whereas, according to the S–R/reinforcement view (Hull, 1943), the association of contiguously active sensory and motor processes is strengthened by a reinforcement-related signal that does not itself form part of the associative structure and acts merely as a catalyst to increase the associative strength between S and R. On this latter view, every time an action is reinforced its association with any prevailing, contiguously active, situational stimuli is increased. This view predicts, naturally enough, that S–R learning should be constrained to the stimuli under which training is conducted and, indeed, in an interesting demonstration, Killcross and Coutureau (2003) found evidence of contextual control of S–R learning induced by overtraining one action in one context and undertraining a different action in a different context. When tested in the overtrained context the action was insensitive to outcome devaluation; when tested in the undertrained context, it was not and showed the normal sensitivity to devaluation indicative of control by an R–O learning process. By extension, the training of actions under distinct discriminative stimuli should also be anticipated to restrict S–R control to the performance assessed under that particular stimulus relative to other stimuli, consistent with aspects of previous reports (Colwill and Rescorla, 1990; De Wit *et al*, 2006).

In contrast, R–O learning is not determined by simple contiguity between action and outcome. This has been demonstrated in a variety of studies, first by Hammond (1980) and later in a number of better-controlled demonstrations (Balleine and Dickinson, 1998a; Colwill and Rescorla, 1986; Dickinson and Mulatero, 1989; Williams, 1989). Thus, for example, when action–outcome contiguity is maintained, non-contiguous delivery of the outcome of an action causes a selective reduction in the performance of that action relative to actions not paired with that outcome. Generally, if a specific outcome is more probable given performance of a specific action, then the strength of the R–O association increases. If the outcome has an equal or greater probability of delivery in the absence of the action, then the R–O association declines. This increase and decrease in the strength of association suggests that, unlike S–R learning, R–O learning is sensitive to the contingency between R and O rather than R–O contiguity. In contrast, S–R learning, being contiguity driven, should be insensitive to non-contiguous outcome delivery. In one study, Dickinson *et al* (1998) found that, compared with relatively under trained rats, overtrained rats were insensitive to a shift in contingency induced by the imposition of an omission schedule; ie, when rats were asked to withhold the performance of an action to earn a sucrose outcome, undertrained rats were able to do so, whereas overtrained rats were not. Thus, reversal of the contingency from positive (O depends on R) to negative (O depends on

withholding R) affected undertrained actions, sensitive to outcome devaluation, but did not affect overtrained actions.

Although the influence of overtraining on contingency sensitivity has not been assessed in human subjects, there is considerable evidence that rat actions and human causal judgments exhibit a comparable sensitivity to the degradation of the action–outcome contingency produced by the delivery of unpaired outcomes. Recall that an instrumental action can be rendered causally ineffective by delivering unpaired outcomes with the same probability as paired ones. In fact, just as the rate of lever pressing by rats declines systematically as the difference in the probability of the paired and unpaired outcomes is reduced (Hammond, 1980), human performance declines similarly when the outcomes are given a nominal value (Shanks and Dickinson, 1991) and, more importantly, so do judgments regarding the causal efficacy of actions (Shanks and Dickinson, 1991; Wasserman et al, 1983).

Summary

These behavioral studies provide consistent evidence of two different associative processes through which actions can be acquired and controlled. One of these, R–O learning, reflects the formation of associations between actions and their consequences or outcomes, a process that explains the sensitivity of newly acquired actions and of choice between distinct courses of action to changes in the value of their consequences and in the contingent relationship between performance of the action and delivery of its associated outcome. Given the clear and substantial regulation by their consequences, the rapid and relatively flexible deployment of actions controlled by the R–O association is clearly consistent with the claim that this learning process is critical to the acquisition and performance of goal-directed action specifically and to executive control of decision-making and choice between courses of action more generally. The second, S–R learning, process involves the formation of an association between the response and antecedent stimuli. Performance controlled by the S–R association is, therefore, stimulus- and, often, contextually bound, and relatively automatic, appearing impulsive or habitual and unregulated by its consequences.

One might anticipate that these distinct learning and behavioral processes would have quite distinct neural determinants and recent research has confirmed this prediction. In the following section, we review evidence suggesting that homologous regions of the cortical-dorsal striatal network are involved in these learning processes in rats and humans, findings that have been established using many of the same behavioral tests described above.

ROLE OF THE CORTICOSTRIATAL NETWORK IN GOAL-DIRECTED AND HABIT LEARNING IN RATS AND HUMANS

Neural Substrates of Goal-Directed Learning

In rats, two components of the corticostriatal circuit in particular have been implicated in goal-directed learning: the prelimbic region of prefrontal cortex—see Figure 1a—and the area of dorsal striatum to which this region of cortex projects: the dorsomedial striatum—Figure 1d (Groenewegen et al, 1990; McGeorge and Faull, 1989; Nauta, 1989). The networks described in this section are illustrated in Figure 2a. Lesions of either of these regions prevents the acquisition of goal-directed learning, rendering performance habitual even during the early stages of training, as assessed using either an outcome devaluation paradigm or contingency degradation (Balleine and Dickinson, 1998a; Corbit and Balleine, 2003b; Yin et al, 2005). Importantly, prelimbic cortex, although necessary for initial acquisition, does not appear to be necessary for the expression of goal-directed behavior; lesions of this area do not impair goal-directed behavior if they are induced after initial training (Ostlund and Balleine, 2005). On the other hand, dorsomedial striatum does appear to be critical for both the learning and expression of goal-directed behavior; lesions of this area impair such behavior if performed either before or after training (Yin et al, 2005).

The finding that the parts of rat prefrontal cortex contribute to action–outcome learning raises the question of whether there exists a homologous region of the primate on prefrontal cortex that contributes to similar functions. A number of fMRI studies in humans have found evidence that a part of the ventromedial prefrontal cortex (vmPFC) is involved in encoding the expected reward attributable to chosen actions, which might suggest this region as a candidate area for a possible homolog. In a typical example, Daw et al (2006) used a four-armed bandit task in which subjects on each trial could choose one of the four bandits to obtain 'points' that could be obtained on the different bandits in varying amounts (that would later be converted into money). The amount of reward expected on a given bandit once it had been chosen was estimated using a reinforcement learning algorithm, which took into account the history of past rewards obtained on that bandit to generate a prediction about the future rewards likely attainable on a given trial. Activity in vmPFC (medial orbitofrontal cortex extending dorsally up the medial wall of prefrontal cortex) was found to correlate with the expected reward value derived from the RL model for the chosen action across trials. A similar finding has been obtained in a number of other studies using similar paradigms and approaches (Hampton et al, 2006; Kim et al, 2006; Tanaka et al, 2004). These findings suggest that human vmPFC is involved in encoding value signals relevant for reward-based action selection; however, the above studies did not deploy the behavioral assays necessary to determine whether such value signals are goal-directed or habitual.

To address this issue, Valentin et al, (2007) performed an outcome devaluation procedure in humans, while subjects were scanned with fMRI during the performance of instrumental actions for food rewards. Following an initial (moderate) training phase, one of these foods was devalued by feeding the subject to satiety on that food. The subjects were then scanned again while they were re-exposed to the instrumental choice procedure in extinction. By testing for
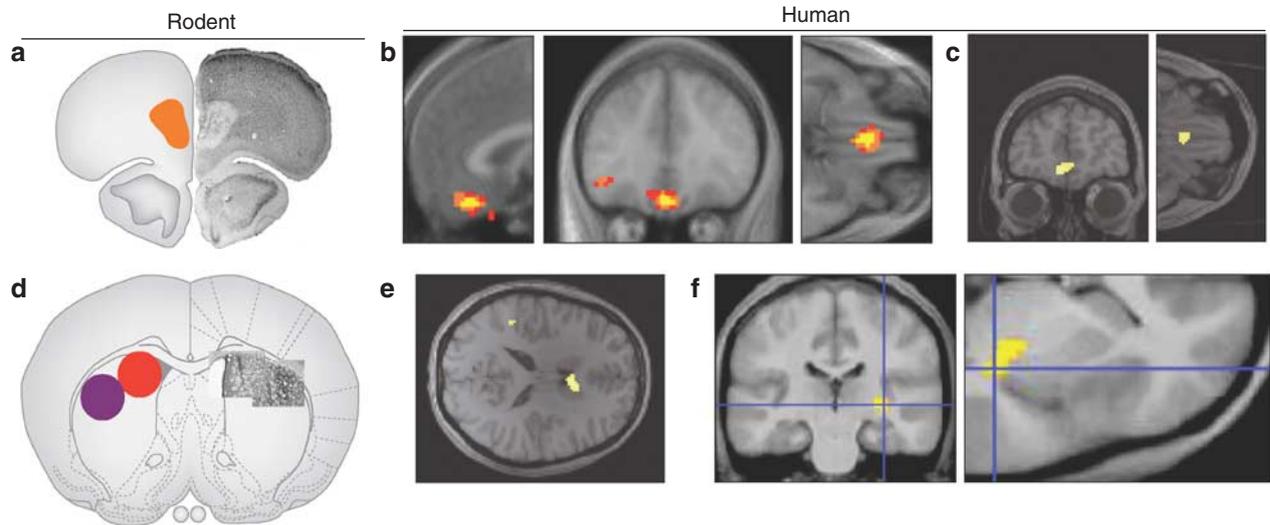
Rodent | Human

Figure 1. (a) Photomicrograph of an NMDA-induced cell body lesion of prelimbic prefrontal cortex (right hemisphere) and approximate region of lesion-induced damage (orange oval; left hemisphere) found to abolish the acquisition of goal-directed action in rats (cf. Balleine and Dickinson, 1998a, b; Corbit and Balleine, 2003a, b; Ostlund and Balleine, 2005). (b) Region of human vmPFC (here medial OFC) exhibiting a response profile consistent with the goal-directed system. Activity in this region during action selection for a liquid food reward was sensitive to the current incentive value of the outcome, decreasing in activity during the selection of an action leading to a food reward devalued through selective satiation compared to an action leading to a non-devalued food reward. From Valentin et al (2007). (c) Regions of human vmPFC (medial prefrontal cortex and medial OFC) exhibiting sensitivity to instrumental contingency and thereby exhibiting response properties consistent with the goal-directed system. Activation plots show areas with increased activity during sessions with a high contingency between responses and rewards compared with sessions with low contingency. From Tanaka et al (2008). (d) Photo-micrographs of NMDA-induced cell-body lesions of dorsomedial and dorsolateral striatum (right hemisphere) with the approximate region of lesion-induced damage illustrated in using red and purple circles, respectively (left hemisphere). This lesion of dorsomedial striatum has been found to abolish acquisition and retention of goal-directed learning (cf. Yin et al, 2005), whereas this lesion of dorsolateral striatum was found to abolish the acquisition of habit learning (Yin et al, 2004). (e) Region of human anterior dorsomedial striatum exhibiting sensitivity to instrumental contingency from the same study described in panel c. (f) Region of posterior lateral striatum (posterior putamen) exhibiting a response profile consistent with the behavioral development of habits in humans. From Tricomi et al, 2009.

regions of the brain showing a change in activity during selection of a devalued action compared with that elicited during selection of a valued action from pre- to post-satiety, it was possible to isolate areas showing sensitivity to the learned action–outcome associations. The regions found to show such a response profile were medial OFC, as well as an additional part of central OFC. The aforementioned results would therefore appear to implicate at least a part of vmPFC in encoding action–outcome and not S–R associations—see Figure 1b. However, another possibility that cannot be ruled out on the basis of the Valentin et al (2007)study alone is that this region may instead contain a representation of the discriminative stimulus (in this case the fractals), used to signal the different available actions, and that the learned associations in this region may be formed between these stimuli and the outcomes obtained. In other words, the design of the Valentin et al (2007) study does not allow us to rule out a possible contribution of vmPFC to purely Pavlovian stimulus-driven valuation signals (see section 'The corticolimbic-ventral striatal network and the motivation of decision making' for further discussion of this point).

Direct evidence of a role for vmPFC in encoding action-related value signals was provided by Gläscher et al (2009). In this study, the possible contribution of discriminative stimuli in driving expected-reward signals in vmPFC was

probed using a specific design in which, in an 'action-based' condition, subjects had to choose between performing one of two different physical motor responses (rolling a trackerball vs pressing a button) in the absence of explicit discriminative stimuli signaling those actions. Subjects were given monetary rewards on a probabilistic basis according to their choice of the different physical actions, and the rewards available on the different actions changed over time. Similar to the results found in studies where both discriminative stimulus information and action-selection components are present, in this task, activity in vmPFC was found to track the expected reward corresponding to the chosen action. These results suggest that activity in vmPFC does not necessarily depend on the presence of discriminative stimuli, indicating that this region contributes to encoding of action-related value signals above and beyond any contribution this region might make to encode stimulus-related value signals.

Further evidence that human vmPFC has a role in goal-directed learning, and in encoding action–outcome-based value signals specifically, has come from a study by Tanaka et al (2008)—see Figure 1c. In this study, rather than using outcome devaluation, areas exhibiting sensitivity to the contingency between actions and outcomes were assessed. As described earlier, sensitivity to action–outcome contingency is another key feature besides sensitivity to
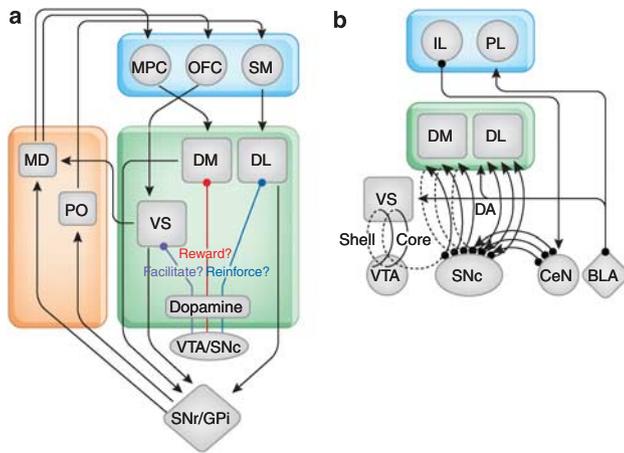
Figure 2. (a) Evidence reviewed in text suggests that distinct neural networks mediate the acquisition of goal-directed actions and habits and the role of goal values and of Pavlovian values in the motivation of performance. On this view, habits are encoded in a network involving sensory-motor (SM) cortical inputs to dorsolateral striatum (DL), with feedback to cortex through substantial nigra reticulata/internal segment of the globus pallidus (SNr/GPi) and posterior thalamus (PO) and are motivated by midbrain dopaminergic inputs from substantia nigra pars compacta (SNc). A parallel circuit linking medial prefrontal cortex (MPC), dorsomedial striatum (DM), SNr, and mediodorsal thalamus (MD) mediates goal-directed actions that may speculatively involve a dopamine-mediated reward process. Finally, choice between actions can be facilitated both by the value of the goal or outcome associated with an action, likely involving amygdala inputs to ventral striatum, MPC and DM, and by Pavlovian values mediated by a parallel ventral circuit mediated by orbitofrontal cortex (OFC) and ventral striatal (VS) inputs into the habit and goal-directed loops. (b) Various theories have been advanced, based on rat and primate data, regarding how limbic, cortical, and midbrain structures interact with the striatum to control performance (see text). Here, dopaminergic (DA) feedforward and feedback processes are illustrated involving VTA-accumbens shell and core and SNc-dorsal striatal networks. The involvement of the BLA in reward processes is illustrated, as is the hypothesized involvement of the infralimbic cortex (IL) and central nucleus of the amygdala in the reinforcement signal derived from SNc afferents on dorsolateral stiatum.

changes in outcome value that distinguishes goal-directed learning from its habitual counterpart. To study this process in humans, Tanaka et al (2008) abandoned the traditional trial-based approach typically used in experiments using humans and nonhuman primates, in which subjects are cued to respond at particular times in a trial, for the unsignaled, self-paced approach more often used in studies of associative learning in rodents in which subjects themselves choose when to respond. Subjects were scanned with fMRI while in different sessions; they responded on four different free operant reinforcement schedules which varied in the degree of contingency between responses shown and rewards obtained. Consistent with the findings from outcome devaluation (Valentin et al, 2007), activity in two subregions of vmPFC (medial OFC and medial prefrontal cortex) as well as one of the target areas of these structures in the human striatum, the anterior caudate nucleus (Haber et al, 2006; Ongür and Price, 2000) was elevated on average across a session when subjects were

performing on a high-contingency schedule compared with when they were performing on a low-contingency schedule —see Figure 1e. Moreover, in the subregion of vmPFC identified on the medial wall, activity was found to vary not only with the overall contingency averaged across a schedule, but also with a locally computed estimate of the contingency between action and outcome that tracks rapid changes in contingency over time within a session, implicating this specific subregion of medial prefrontal cortex in the on-line computation of contingency between actions and outcomes. Finally, activation of medial prefrontal cortex also tracked a measure of subjective contingency; ie, the ratings of the subjects regarding the causal efficacy of their actions. This rating, taken after each trial block, positively correlated (approximately 0.6) with measures of objective contingency, suggesting that the medial vmPFC-caudate network may interact directly with medial prefrontal cortex to influence causal knowledge.

Neural Substrates of Habit Learning

The finding that medial prefrontal cortex and its striatal efferents contribute to goal-directed learning in both rats and humans, raises the question as to where in the corticostriatal network habitual S–R learning processes are implemented. Considerable earlier, although behaviorally indirect, evidence from studies using tasks that are nominally procedural and could potentially involve S–R learning (largely simple skill learning in humans or maze learning in rats) has implicated a region of dorsal striatum lateral to the caudate nucleus—referred to as dorsolateral striatum in rat or putamen in primates—in S–R encoding (see Figure 2a). More direct evidence, based on the criteria described in section 'Multiple sources of action control in rodents and humans: goals and habits,' was provided in a study by Yin et al (2004). Rats with lesions to a region of dorsolateral striatum were found to remain goal-directed even after extensive training which, in sham-lesioned controls, led to clear habitization; ie, whereas actions in lesioned rats remained sensitive to outcome devaluation those of sham controls did not. This increased sensitivity to the consequences of actions was observed both with outcome devaluation and contingency degradation procedures; in the latter case, overtrained rats were unable to adjust their performance of an action when responding caused the omission of reward delivery, whereas inactivation of dorsolateral striatum rendered rats sensitive to this omission contingency (Yin et al, 2006). This finding suggests that this region of dorsolateral striatum has a critical role in the habitual control of behavior in rodents—see Figure 1d.

To establish whether a similar area of striatum also contributes to such a process in humans, Tricomi et al (2009) scanned subjects with fMRI while they performed on a variable interval schedule for food rewards in which one group of subjects was overtrained to induce behavioral habitization. In the group that was given this procedure,

activity in a region of lateral striatum (caudo-ventral putamen) was found to show an increased activation on the third day of training when an outcome devaluation test revealed subjects' responding to be habitual, compared with the first day of training when responding in undertrained subjects was shown to be goal-directed—see Figure 1f. These findings provide evidence to suggest that this region of posterolateral putamen in humans may correspond functionally to the area of striatum found to be critical for habitual control in rodents. Additional hints of a role for human caudoventral striatum in habitual control can be gleaned from fMRI studies of 'procedural' sequence learning (Jueptner et al, 1997a; Lehéricy et al, 2005). Such studies have reported a transfer of activity within striatum from the anterior striatum to posterior striatum as a function of training. Although these earlier studies did not formally assess whether behavior was habitual by the time that activity in posterolateral striatum had emerged, these studies did show that, by this time, sequence generation was insensitive to dual-task interference, a behavioral manipulation potentially consistent with habitization.

Summary

The evidence reviewed in this section, summarized in Figure 1, suggests that there is considerable overlap in the corticostriatal circuitry that mediates goal-directed and habitual actions in humans and rodents. Rodent studies have implicated prelimbic cortex and its striatal efferents on dorsomedial striatum as a key circuit responsible for goal-directed learning. In a series of fMRI studies, vmPFC has been found to be involved in encoding reward predictions based on goal-directed action–outcome associations in humans, suggesting that this region of cortex in the primate prefrontal cortex is a likely functional homolog of prelimbic cortex in the rat. Furthermore, the area of anterior caudate nucleus found in humans to be modulated by contingency would seem to be a candidate homolog for the region of dorsomedial striatum implicated in goal-directed control in the rat. Finally, the evidence reviewed here supports the suggestion that a region of dorsolateral striatum in rodents and of the putamen in humans is involved in the habitual control of behavior, which when taken together with the findings on goal-directed learning reviewed previously, provides converging evidence that the neural substrates of these two systems for behavioral control are relatively conserved across mammalian species.

THE CORTICOLIMBIC-VENTRAL STRIATAL NETWORK AND THE MOTIVATION OF DECISION MAKING

The findings described above provide evidence of distinct sources of action control governed by R–O and S–R learning processes. Given the differences in associative structure and

in the neural systems controlling these types of learning, it should not be surprising to discover that they are also governed by different motivational processes. Generally, these processes can be differentiated into the influence of the encoded reward-value of the outcome of an action, derived from direct consummatory experience, and the influenced of Pavlovian predictors of reward. Experienced reward determines the performance of goal-directed actions and, hence, reflects the control of actions by outcome values. Where there is a significant degree of stimulus control over action selection, however, it is possible to observe the influence of Pavlovian stimuli associated with reward on performance; ie, stimuli that predict rewarding events can enhance the influence of action selection resulting in an increased tendency to perform selected actions and with increased vigor. As such, this motivational influence is referred to here as that based on Pavlovian values.

Neural Basis of Outcome Values

It is now well established that the reward processes that establish outcome values depend on the ability of animals to evaluate the affective and motivationally relevant properties of the goal or outcome of goal-directed actions (Balleine, 2001, 2004; Dickinson and Balleine, 1994, 2002). It is one of more striking properties of goal-directed actions that performance is determined by the experienced reward value of the outcomes associated with actions and, unlike habits or other reflexes, is not directly affected by shifts in primary motivation (Dickinson and Balleine, 1994). For example, rodents do not immediately alter their choice of an action associated with a more (or less) calorific outcome when deprivation is increased (or decreased) and do so only after they have experienced the outcome value in that new deprivation state (Balleine, 1992; Balleine et al, 1995; Balleine and Dickinson, 1994). The influence of this form of reward or incentive learning on goal-directed actions has been reported following wide ranging shifts in motivational state induced by: (i) specific satiety-induced outcome devaluation (Balleine and Dickinson, 1998b); (ii) taste aversion-induced outcome devaluation (Balleine and Dickinson, 1991); (iii) shifts from water deprivation to satiety (Lopez et al, 1992); (iv) changes in outcome value mediated by drug states and withdrawal (Balleine et al, 1994; Balleine et al, 1995; Hellemans et al, 2006; Hutcheson et al, 2001); and, (v) changes in the value of thermoregulatory reward (Hendersen and Graham, 1979); and (vi) sexual reward (Everitt and Stacey, 1987; Woodson and Balleine, 2002) (see Balleine, 2001; Balleine, 2004; Balleine, 2005; Dickinson and Balleine, 2002 for reviews).

Current theories suggest, therefore, that outcome values are established by associating the specific sensory features of outcomes with emotional feedback (Balleine, 2001; Dickinson and Balleine, 2002). Given these theories, one might anticipate that neural structures implicated in associations of this kind would have a critical role in goal-directed action. The amygdala, particularly its baso-

lateral region (BLA), has long been argued to mediate sensory-emotional association, and recent research has established the involvement of this area in goal-directed action in rodents. The BLA has itself been heavily implicated in a variety of learning paradigms that have an incentive component (Balleine and Killcross, 2006); for example, this structure has long been thought to be critical for fear conditioning and has recently been reported to be involved in a variety of feeding-related effects, including sensory-specific satiety (Malkova et al, 1997), the control of food-related actions (see below), and in food consumption elicited by stimuli associated with food delivery (Holland et al, 2002; Petrovich et al, 2002). And, indeed, in several recent series of experiments clear evidence has emerged for the involvement of BLA in incentive learning. In one series, we found that lesions of the BLA rendered the instrumental performance of rats insensitive to outcome devaluation, apparently because they were no longer able to associate the sensory features of the instrumental outcome with its incentive value (Balleine et al, 2003; Corbit and Balleine, 2005). This suggestion was confirmed using post-training infusions of the protein-synthesis inhibitor, anisomycin, after exposure to an outcome after a shift in primary motivation. In this study, evidence was found to suggest that the anisomycin infusion blocked both the consolidation and the reconsolidation of the stimulus--affect association underlying incentive learning (Wang et al, 2005). More recently, we have found direct evidence for the involvement of an opioid receptor-related process in the basolateral amygdala in encoding outcome value based on sensory–affect association. In this study, infusion of naloxone into the BLA blocked the assignment of an increase in the value of a sugar solution when it was consumed in an increased state of food deprivation without affecting palatability reactions to the sucrose. Naloxone infused into the ventral pallidum or accumbens shell had the opposite effect, reducing palatability reactions without affecting outcome value (Wassum et al, 2009).

With respect to the encoding of outcome values, the effects of amygdala manipulations on feeding have been found to involve connections between the amygdala and the hypothalamus (Petrovich et al, 2002) and, indeed, it has been reported that neuronal activity in the hypothalamus is primarily modulated by chemical signals associated with food deprivation and food ingestion, including various macronutrients (Levin, 1999; Seeley et al, 1996; Wang et al, 2004; Woods et al, 2000). Conversely, through its connections with visceral brain stem, midline thalamic nuclei, and associated cortical areas, the hypothalamus is itself in a position to modulate motivational and nascent affective inputs into the amygdala. Together with the findings described above, these inputs, when combined with the amygdala's sensory afferents, provide the basis for a simple feedback circuit linking this sensory information with motivation/affective feedback to determine outcome value. Nevertheless, it is not clear from this structural perspective how changes in outcome value encoded on the basis of this feedback act to influence the selection and initiation of specific courses of action.

The BLA projects to a variety of structures in the cortico-basal ganglia network implicated in the control of goal-directed action, such as the prelimbic cortex, mediodorsal thalamus, and dorsomedial striatum (see Figure 2b). However, the role of the BLA and these associated regions in action control differ in important ways (Ostlund and Balleine, 2005; Ostlund and Balleine, 2008) raising the question of how the outcome values established in the BLA make contact with the basal ganglia network critical for goal-directed motor control. Evidence suggests that BLA activity is necessary for encoding outcome values in insular cortex, particularly its gustatory region (Balleine and Dickinson, 2000), and that these values are then distributed to regions of prefrontal cortex and striatum to control action (Condé et al, 1995; Rodgers et al, 2008). With regard to the latter, insular cortex projects to the ventral striatum, particularly to the core of the nucleus accumbens (NACco) (Brog et al, 1993) and lesions of the NACco have been found (i) to impair instrumental performance (Balleine and Killcross, 1994) and, in contrast to more medial regions, such as the shell and central pole (de Borchgrave et al, 2002), (ii) to reduce sensitivity to outcome devaluation (Corbit et al, 2001). However, these lesions have no effect on the sensitivity of rats to the degradation of the instrumental contingency (Corbit et al, 2001), suggesting that the NACco influences goal-directed performance, but does not affect goal-directed learning. Generally, therefore, the NACco, as a component of the limbic cortico-basal ganglia network, appears to mediate the ability of the incentive value of rewards to affect instrumental performance, but does not have a direct role in action–outcome learning per se (Balleine and Killcross, 1994; Cardinal et al, 2002; Corbit et al, 2001; de Borchgrave et al, 2002; Parkinson et al, 2000). Thus, as has long been argued, the NACco appears to have a central role in the translation of motivation into action by bringing changes in outcome value to bear on performance consistent with its description as the limbic–motor interface (Mogenson et al, 1980).

In humans, the evidence on the role of the amygdala in outcome valuation is somewhat ambiguous, although broadly compatible with the aforementioned evidence from the rodent literature. Although some studies have reported amygdala activation in response to the receipt of rewarding outcomes, such as pleasant tastes or monetary reward (Elliott et al, 2003; O'Doherty et al, 2003; O'Doherty et al, 2001a; O'Doherty et al, 2001b), other studies have suggested that the amygdala is more sensitive to the intensity of a stimulus rather than its value (Anderson et al, 2003; Small et al, 2003), as amygdala responds equally to both positive and negative valenced stimuli matched for intensity. These latter findings could suggest a more general role for the amygdala in arousal rather than valuation per se, although, alternatively, the findings are also compatible with the possibility that both positive and negative outcome valuation signals are present in the amygdala (correlating both

positively and negatively with outcome values, respectively), and that such signals are spatially intermixed at the single neuron level (Paton *et al*, 2006). Indeed in a followup fMRI study by Winston *et al* (2005), BOLD responses in amygdala were found to be driven best by an interaction between valence and intensity (that is by stimulus of high intensity and with high valence), rather than by one or other dimension alone, suggesting a role for this region in the overall value assigned to an outcome, which would be a product of its intensity (or magnitude) and its valence.

Even clearer evidence for the presence of outcome valuation signals has been found in human vmPFC (particularly in medial orbitofrontal cortex) and adjacent central orbitofrontal cortex. Specifically, activity in medial orbitofrontal cortex correlates with the magnitude of monetary outcome received (O'Doherty *et al*, 2001a), and medial along with central orbitofrontal cortex correlates with the pleasantness of the flavor or odor of a food stimulus (Kringelbach *et al*, 2003; Rolls *et al*, 2003). Furthermore, activity in these regions decreases as the hedonic value of that stimulus decreases as subjects are sated on it (Kringelbach *et al*, 2003; O'Doherty *et al*, 2000; Small *et al*, 2003). De Araujo *et al* (2003) found that activity in caudal orbitofrontal cortex correlated with the subjective pleasantness of water in thirsty subjects, and, moreover, that insular cortex was active during the receipt of water when subjects were thirsty compared with when they were sated, suggesting the additional possible involvement of at least a part of insular cortex in some features of outcome valuation in humans. Further evidence of a role for medial orbitofrontal cortex in encoding the values of goals has come from a study by Plassmann *et al* (2007) who used an economic auction mechanism to elicit subjects' subjective monetary valuations for different goal objects, which were pictures of food items one of which subjects would later have the opportunity to consume depending on their assigned valuations. Activity in medial orbitofrontal cortex was found to correlate with subjective valuations for the different food items. Although in this case such value signals are at the time of choice and hence reflect 'goal-values' rather than outcome values, these findings are consistent with a contribution for this region in encoding the value of goals even if those goals are not currently being experienced (as would be required for a region linking action and outcome representations at the time of choice).

Neural Basis of Pavlovian Values

In contrast to goal-directed actions, responses controlled by S–R learning are directly affected by shifts in primary motivation. For example, Dickinson *et al* (1995) found that, when relatively undertrained to lever press for food reward, rats only altered lever-press performance when sated after they had experienced the change in outcome value in that state. In contrast, when overtrained, their performance was reduced directly by satiety suggesting that the vigor of habitual actions is more dependent on the activational

effects of motivational state than undertrained actions. In addition, Holland (2004) found that, as overtrained actions became less sensitive to changes in outcome value, they become increasingly sensitive to the effects of Pavlovian reward-related stimuli on response vigor. This dissociation suggests that the influence of outcome values and of Pavlovian values provide distinct sources of motivation for the performance of actions, the former controlling the performance of goal-directed actions and the latter actions that are under significant stimulus control, whether they have simply been trained under discriminative stimuli or have been overtrained and become habitual. Two other behavioral observations support this argument. First, we have found that, when actions are performed in a chain, performance of the most distal action is controlled by the outcome value and not by Pavlovian values, whereas the most proximal action is motivated primarily by Pavlovian values and not by outcome values (Corbit and Balleine, 2003a). Second, Rescorla (1994)—and later Holland (2004)—found that the influence of a Pavlovian stimulus on lever-press responding was not affected by earlier devaluation of the outcome associated with that stimulus (see also Balleine and Ostlund, 2007d). Hence, reward-related stimuli affect the performance of instrumental actions irrespective of the value of the outcome that the stimulus predicts, an effect consistent with the argument that Pavlovian values exert their effects on actions through stimulus, as opposed to outcome, control. Indeed, this same insensitivity to outcome devaluation has been reported for actions that are acquired and maintained not by primary reward but by stimuli previously associated with reward; ie, so called conditioned reinforcers. Thus, lever pressing trained with a conditioned reinforcer is maintained even when the outcome associated with the stimulus earned by the action has been devalued (Parkinson *et al*, 2005).

The observation by Rescorla (1994) was particularly important because it provided a very clear demonstration of the fact that, not only the general excitatory influence but also very specific effects of reward-related stimuli on action selection are unaffected by outcome devaluation. This latter, specific effect of stimuli on action selection (an effect referred to as outcome-specific Pavlovian-instrumental transfer, or simply specific transfer) is demonstrated when two actions trained with different outcomes are performed in the presence of a stimulus paired with one or other of those outcomes; in this situation only the action that earned the same outcome as the stimulus is influenced by that cue. It is also important to note here that we have also recently established evidence of a comparable effect in human subjects. In similar manner to findings in rats, a Pavlovian stimulus associated with a specific outcome (eg, orange juice, chocolate milk, or coca cola) only elevated the performance of actions that earned the outcome predicted by that stimulus and did not affect actions trained with other outcomes (Bray *et al*, 2008).

There are two important aspects of the neural bases of Pavlovian values to consider relating to (1) the neural bases

of predictive learning generally and (2) the neural processes through which predictive learning influences decision making. These are illustrated in Figure 2a and b.

With regard to the former, using maniulations of the Pavlovian stimulus–outcome contingency, Ostlund and Balleine (2007b) found that the orbitofrontal cortex in rats has an important role in establishing the predictive validity of Pavlovian cues with regard to specific outcomes. Thus, in control rats, although conditioned responding to a cue declined when its outcome was also delivered unpaired with that cue, responding to a stimulus predicting a different outcome was unaffected. In contrast, rats with lesions of lateral orbitofrontal cortex reduced responding to both stimuli suggesting that these animals were unable to discriminate the relative predictive status of cues with respect to their specific outcomes. Likewise, single-unit recording studies in both rodents and nonhuman primates have implicated neurons in a network of brain regions including the orbitofrontal cortex, and its striatal target area, the ventral striatum, notably the accumbens shell (Fudge and Emiliano, 2003; Haber, 2003), in encoding stimulus–reward associations. For example, Schoenbaum et al (1998) used a go/no-go reversal task in rats in which, on each trial, one of two different odor cues signaled whether or not a subsequent nose poke by the rat in a food well would result in the delivery of an appetitive sucrose solution or an aversive quinine solution. Neurons in the orbitofrontal cortex were found to discriminate between cues associated with the positive and negative outcomes, and some were also found to show an anticipatory response related to the expected outcome after the animal had placed its head in the food well, but immediately before the outcome was delivered. Cue-related anticipatory responses have also been found in orbitofrontal cortex relating to the behavioral preference of monkeys for a predicted outcome; ie, the responses of the neurons to the cue paired with a particular outcome depend on the relative preference of the monkey for that outcome compared with another outcome presented in the same block of trials (Tremblay and Schultz, 1999).

Furthermore, bilateral lesions of orbitofrontal cortex, or crossed unilateral lesions of the orbitofrontal cortex in one hemisphere with the amygdala (a region connecting strongly to both the orbitofrontal cortex and the ventral striatum (Alheid, 2003; Ongür and Price, 2000)) in the other hemisphere, result in impairments in the modulation of conditioned Pavlovian responses following changes in the value of the associated outcome induced by an outcome devaluation procedure in both rats and monkeys (Baxter et al, 2000; Hatfield et al, 1996; Malkova et al, 1997; Ostlund and Balleine, 2007b; Pickens et al, 2003). Perhaps unsurprisingly given the connectivity between OFC and ventral striatum, activity of some neurons in this region have, similar to those in the orbitofrontal cortex, been found to reflect expected reward in relation to the onset of a stimulus presentation, and to track progression through a task sequence ultimately leading to reward (Cromwell and

Schultz, 2003; Day et al, 2006; Shidara et al, 1998). Some studies report that lesions of a part of the ventral striatum, the nucleus accumbens core can impair learning or expression of Pavlovian approach behavior (Parkinson et al, 1999), or that infusion of a dopaminergic agonist into accumbens can also impair such learning (Parkinson et al, 2002), suggesting an important role for ventral striatum in facilitating the production of at least some classes of Pavlovian conditioned associations.

Consistent with these findings in rodents and nonhuman primates, functional neuroimaging studies in humans have also revealed activations in both of these areas during both appetitive and aversive Pavlovian conditioning in response to Pavlovian cue presentation. For example, Gottfried et al (2002) reported activity in orbitofrontal cortex and ventral striatum (as well as the amygdala) following presentation of visual stimuli predictive of the subsequent delivery of both a pleasant and an unpleasant odor. Gottfried et al (2003) aimed at investigating the nature of such predictive representations to establish whether these responses were related to the sensory properties of the unconditioned stimulus irrespective of its underlying reward value, or whether they were directly related to the reward value of the associated unconditioned stimulus. To address this, Gottfried et al (2003) trained subjects to associate visual stimuli with one of two food odors: vanilla and peanut butter while being scanned with fMRI. Subjects were then removed from the scanner and fed to satiety on a food corresponding to one of the food odors to devalue that odor, similar to the devaluation procedures described earlier during studies of instrumental action selection. Following the devaluation procedure, subjects were then placed back in the scanner and presented with the conditioned stimuli again in a further conditioning session. Neural responses to presentation of the stimulus paired with the devalued odor were found to decrease in orbitofrontal cortex and ventral striatum (and amygdala) from before to after the satiation procedure, whereas no such decrease was evident for the stimulus paired with the non-devalued odor. These results suggest that the orbitofrontal cortex and ventral striatum encode both the value of event predicted by a stimulus and not merely its sensory features.

Learning Pavlovian Values

The finding of expected value signals in the brain raises the question of how such signals are learned in the first place. An influential theory by Rescorla and Wagner (1972) suggests that the learning of Pavlovian predictions is mediated by the degree of surprise engendered when an outcome is presented, or more precisely the difference between what is expected and what is received. Formally this is called a prediction error, which in the Rescorla–Wagner formulation can take on either a positive or negative sign depending on whether an outcome is greater than expected (which would lead to a positive error signal), or less than expected (which would lead to a negative error).

This prediction error is then used to update predictions associated with a particular stimulus or cue in the environment, so that if this cue always precedes, for eaxmple, a reward (and hence is fully predictive of reward) eventually the expected value of the cue will converge to the value of the reward, at which point the prediction error is zero and no further learning will take place.

Initial evidence for prediction error signals in the brain emerged from the work of Wolfram Schultz and colleagues who observed such signals by recording from the phasic activity of dopamine neurons in awake behaving nonhuman primates undergoing simple Pavlovian or instrumental conditioning tasks with reward (Hollerman and Schultz, 1998; Mirenowicz and Schultz, 1994; Schultz, 1998; Schultz et al, 1997). These neurons, particularly those present in the ventral tegmental area in the midbrain, project strongly to the ventral corticostriatal circuit, including the ventral striatum and orbitofrontal cortex (Oades and Halliday, 1987). The response profile of these dopamine neurons closely resembles a specific form of prediction error derived from the temporal difference learning rule, in which predictions of future reward are computed at each time interval within a trial, and the error signal is generated by computing the difference in successive predictions (Montague et al, 1996; Schultz et al, 1997). Similar to the temporal difference prediction error signal, these neurons increase their firing when a reward is presented unexpectedly; decrease their firing from baseline when a reward is unexpectedly omitted; respond initially at the time of the US before learning is established but shift back in time within a trial to respond instead at the time of presentation of the CS once learning has taken place. Further evidence in support of this hypothesis has come from recent studies using fast cyclical voltammetry assays of dopamine release in the ventral striatum during Pavlovian reward conditioning, in which timing of dopamine release in ventral striatum was also found to exhibit a shifting profile, occurring initially at the time of reward presentation but gradually shifting back to occur at the time of presentation of the reward predicting cue (Day and Carelli, 2007).

To test for evidence of a temporal difference prediction error signal in the human brain, O'Doherty et al (2003) scanned human subjects while they underwent a classical conditioning paradigm in which associations were learned between arbitrary visual fractal stimuli and a pleasant sweet taste reward (glucose). The specific trial history that each subject experienced was next fed into a temporal difference model to generate a time series that specified the model-predicted prediction error signal that was then regressed against the fMRI data for each individual subject to identify brain regions correlating with the model-predicted time series. This analysis revealed significant correlations with the model-based predictions in a number of brain regions, most notably in the ventral striatum (ventral putamen bilaterally) as well as weaker correlations with this signal in orbitofrontal cortex. Consistent with these findings, numerous other studies have found evidence implicating ventral

striatum in encoding prediction errors (Abler et al, 2006; McClure et al, 2003; O'Doherty et al, 2004). In a recent study by Hare et al (2008), the role of ventral striatum and orbitofrontal cortex in encoding prediction errors for free rewards was tested against a number of other types of valuation related-signals while subjects performed a simple decision-making task. Activity in ventral striatum was found to correlate specifically with reward prediction errors, whereas activity in medial orbitofrontal cortex was found to correspond more to the valuation of the goal of the decision, perhaps consistent with the aforementioned role for this region in goal-directed learning.

The studies discussed above demonstrate that prediction error signals are present during learning of Pavlovian stimulus–reward associations, a finding consistent with the tenets of a prediction error-based account of associative learning. However, merely demonstrating the presence of such signals in either dopamine neurons or in target areas of these neurons, such as the striatum, during learning does not establish whether these signals are causally related to learning or merely an epi-phenomenon. The first study in humans aiming to uncover a causal link was that of Pessiglione et al (2006) who manipulated systemic dopamine levels by delivering a dopamine agonist and antagonist while subjects were being scanned with fMRI during performance of a reward–learning task. Prediction error signals in striatum were boosted following administration of the dopaminergic agonist, and diminished following administration of the dopaminergic antagonist. Moreover, behavioral performance followed the changes in striatal activity and was increased following administration of the dopamine agonist and decreased following administration of the antagonist. These findings support, therefore, a causal link between prediction error activity in striatum and the degree of Pavlovian conditioning.

The Influence of Pavlovian Values on Decision Making

Over and above the role of ventral striatum in acquiring Pavlovian predictions, this region has also been implicated in the way Pavlovian values alter choice, particularly in studies of Pavlovian-instrumental transfer. In both rodents and humans, however, it appears not to involve the accumbens core directly but the surrounding accumbens shell/ventral putamen region of ventral striatum. In rats, considerable evidence suggests that the amygdala has a role both in Pavlovian conditioning proper (Ostlund and Balleine, 2008) but also in the way that Pavlovian cues influence both the vigor of responses, involving the central nucleus of the amygdala (Hall et al, 2001; Holland and Gallagher, 2003), and in the way that these cues influence choice, the latter dependent on basolateral amygdala. In one study, Corbit and Balleine (2005) doubly dissociated the influence of cues on choice and on responses vigor; lesions of amygdala central nucleus abolished the increase in vigor induced by these cues without affecting their biasing

influence on choice. In contrast, lesions of basolateral amygdala abolished the effect on choice without affecting the increase in vigor. In humans, evidence has emerged linking the degree of activity in the right amygdala across subjects with the extent to which a Pavlovian conditioned cue modulates the vigor of responding (Talmi *et al*, 2008). However, as yet, no study has assessed the functions of different amygdalar subnuclei on the motivational control of behavior as has been carried out in rodents.

Circuitry involving amygdala has recently been implicated in these effects of Pavlovian values on choice. The basolateral amygdala projects to mediodorsal thalamus (Reardon and Mitrofanis, 2000), to orbitofrontal cortex, and to both the core and shell regions of the nucleus accumbens (Alheid, 2003) and, in fact lesions of each of these regions appear to influence outcome specific form of Pavlovian-instrumental transfer (Ostlund and Balleine, 2005; Ostlund and Balleine, 2007b; Ostlund and Balleine, 2008) in subtly different ways. Lesions of mediodorsal thalamus and of orbitofrontal cortex have similar effects. When rats were trained on two actions for different outcomes, a stimulus associated with one of the outcomes is usually observed to bias choice and to elevate only the action trained with the outcome predicted by that stimulus. In lesioned rats, however, the stimuli were found to elevate performance indiscriminately; ie, lesioned rats increased their performance but of both actions equally (Ostlund and Balleine, 2007b; Ostlund and Balleine, 2008). In contrast, lesions of ventral striatum have a very different effect. Although lesions of the accumbens core have been reported to affect the vigor of responses in the presence of Pavlovian cues, Corbit *et al* (2001) found that these lesions had no effect on the biasing effects on choice performance. Thus, rats with lesions of the accumbens core that were trained on two actions for different outcomes were just as sensitive as sham lesioned rats to the effects of a stimulus associated with one of the two outcomes; ie, similar to shams they increased responding of the action that in training earned the outcome predicted by the stimulus and did not alter performance of the other action. In contrast, lesions of the accumbens shell completely abolished outcome selective Pavlovian-instrumental transfer. Although other evidence suggested that the rats learned both the Pavlovian S–O and instrumental R–O relationships to which they were exposed, the Pavlovian values predicted by the stimuli had no effect on either the vigor of responding or on their choice between actions.

In humans, Talmi *et al* (2008) reported that BOLD activity in the central nucleus accumbens (perhaps analogous to the core region in rodents) was engaged when subjects' were presented with a reward-predicting Pavlovian cue while performing an instrumental response, which led to an increase in the vigor of responding, consistent with the effects of general Pavlovian to instrumental transfer. In a study of outcome-specific Pavlovian-instrumental transfer in humans using fMRI, Bray *et al* (2008) trained subjects on instrumental actions each leading to one of four different

unique outcomes. In a separate Pavlovian training session, subjects were previously trained to associate different visual stimuli with the subsequent delivery of one of these outcomes. Specific transfer was then assessed by inviting subjects to choose between pairs of instrumental actions that, in training, were associated with the different outcomes in the presence of a Pavlovian visual cue that predicted with one of those outcomes. Consistent with the effects of specific transfer, subjects were biased in their choice toward the action leading to the outcome consistent with that predicted by the Pavlovian stimulus. In contrast to the region of accumbens activated in the general transfer design of Talmi *et al* (2008), specific transfer produced BOLD activity in a region of ventrolateral putamen: this region was less active in trials in which subjects chose the action incompatible with the Pavlovian cue compared with trials in which they chose compatible action, or indeed in other trials in which a Pavlovian stimulus paired with neither outcome was presented. These findings could suggest a role for this ventrolateral putamen region in linking specific outcome–response associations with Pavlovian cues and suggest that on occasions when an incompatible action is chosen, activity in this region may be inhibited. Given the role of this more lateral aspect of the ventral part of the striatum in humans in specific-PIT, it might be tempting to draw parallels between the functions of this area in humans with that of the shell of the accumbens implicated in specific-transfer in rodents. At the moment such suggestions must remain speculative until more fine-grained studies of this effect are conducted in humans, perhaps making use of higher resolution imaging protocols to better differentiate between different ventral striatal (and indeed amygdala) subregions.

THE INTEGRATION OF ACTION CONTROL AND MOTIVATION IN DECISION MAKING

How learning and motivational processes are integrated to guide performance is a critical issue that has received relatively little attention to date. Some speculative theories have been advanced within the behavioral, computational, and neuroscience literatures, although there is not at present a significant body of experimental results to decide how this integration is achieved.

Behavioral Studies

Although the fact that both the outcome itself and stimuli that predict that outcome can exert a motivational influence on performance suggests that these processes exert a complimentary influence on action selection, the growing behavioral and neural evidence that these sources of motivation are independent, reviewed above, suggests that the behavioral complementarity of these processes is achieved through an integrative network rather than through a common representational or associative process. This kind of conclusion has in fact long been the source of a

number of two-process theories of instrumental conditioning that have been advanced in the behavioral literature. These positions have typically argued for the integration of some form of Pavlovian and instrumental learning process. Although this is not the place to review the variety of theories advanced along these lines (see Balleine and Ostlund, 2007d), the evidence reviewed above, arguing for distinct learning and motivational processes mediating goal-directed and habitual actions, suggests that this may be based on integration across the S–R and R–O domain in the course of normal performance. Although, this may be achieved in variety of ways, some recent evidence suggests that it is accomplished through a distributed outcome representation involving, on the one hand, the outcome as a goal and on the other the outcome as a stimulus with which actions can become associated. That outcomes might have this kind of dual role is particularly likely in self-paced situations where actions and outcomes occur intermittently, sometimes preceding sometimes following one another over time. Thus, in experiments in which food outcomes are used both as discriminative stimuli and as goals, the delivery of the food can be arranged such that it selects (always precedes) one action and independently serves as an outcome of (always follows) a different action. Devaluation of that outcome was found to influence the action with which it was associated as a goal but did not affect the selection of the action for which the outcome served as a discriminative stimulus (De Wit et al, 2006; Dickinson and de Wit, 2003). This kind of finding encourages the view that, in the ordinary course of events, stimuli and goals exert complementary control over action selection and initiation, respectively.

This position is summarized in Figure 3. Here, the outcome controls actions in two ways: (1) through a form of S–R association in which the stimulus properties of the outcome can select an action with which they are associated, ie, through an O^S–R association; and (2) through the standard R–O association in which any selected action retrieves the outcome as a goal; ie, an R–O^G. Thus, given this form of O–R association, anything that retrieves the outcome, such as the Pavlovian S–O association in specific transfer experiments, for example, should act to select the R with which the O is associated (as a stimulus). Selection of that specific R then acts to retrieve the O with which the R is associated for evaluation, now as a goal, allowing the O^S and O^G to work together to control selection and initiation of the action in performance. That something like this is going on is supported by much of the behavioral evidence already discussed above. Thus, outcome devaluation affects neither the ability of discriminative cues or Pavlovian cues to affect performance and only reduces actions that are dependent on outcome value for initiation (Colwill and Rescorla, 1990; Holland, 2004; Rescorla, 1994). Likewise, action selection inducted by reinstatement of an action by outcome delivery after a period of extinction is unaffected by devaluation of the outcome, although the rate of subsequent performance is significantly reduced suggesting that the O^S selection

process is initially engaged normally and the evaluative O^G process only later acts to modify the rate of performance (Balleine and Ostlund, 2007d; Ostlund and Balleine, 2007a).

Furthermore, this account provides a very natural explanation of the acquisition of habits and their dependence on overtraining. Thus, if an O^S process usually exerts control over action selection, then it requires only the assumption that the strength of this selection process grows sufficiently strong with over training to initiate actions without the need for the evaluation or, more likely, before evaluation is completed. Thus, habitual actions have the quality of being impulsive; they are implemented before the actor has had time adequately to evaluate their consequences. On this view, therefore, the Pavlovian and outcome-related motivational processes associated with habits and actions are intimately related, although their joint influence over the final common path to action. Nevertheless, although these data and this general theoretical approach points to a simple integrative process, it does not provide much in the way of constraints as to how such a process is implemented. Several views of this have, however, been developed in both the computational and neuroscience literatures.

Computational Theories of Action Selection

A particularly influential class of models for instrumental conditioning has arisen from a subfield of computational theories collectively known as reinforcement learning (Sutton and Barto, 1998). The core feature of such RL models is that, to choose optimally between different actions, an agent needs to maintain internal representations of the expected reward available on each action, and then subsequently choose the action with the highest expected value. Also central to these algorithms is the notion of a prediction error signal that is used to learn and update expected values for each action through experience, just as was the case for Pavlovian conditioning described earlier. In one such model—the actor/critic, action selection is conceived as involving two distinct components: a critic, which learns to predict future reward associated with particular states in the environment, and an actor which chooses specific actions to move the agent from state to state according to a learned policy (Barto, 1992, 1995). The critic encodes the value of particular states in the world and as such has the characteristics of a Pavlovian reward prediction signal described above. The actor stores a set of probabilities for each action in each state of the world, and chooses actions according to those probabilities. The goal of the model is to modify the policy stored in the actor, such that, over time, those actions associated with the highest predicted reward are selected more often. This is accomplished by means of the aforementioned prediction error signal that computes the difference in predicted reward as the agent moves from state to state. This signal is then used to update value predictions stored in the critic for each state, but also to update action probabilities stored in the
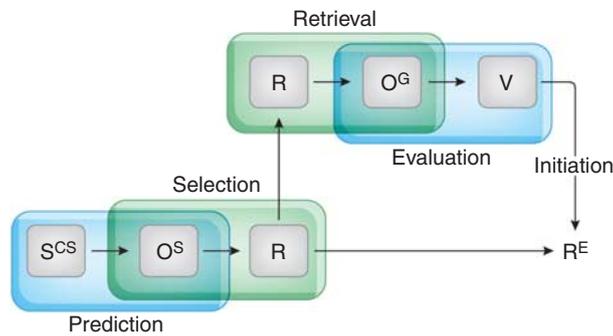
Figure 3. Associations that current evidence suggests are formed between various stimuli (S), actions (R), and outcomes (O) or goals (G) and goal values (V) during the course of acquisition or performance of goal-directed action. (i) Current research provides evidence for both $R–O^G$ and $O^S–R$ associations in the control of performance. Whereas the $O^S–R$ association does not directly engage or influence changes in outcome value, the $R–O^G$ association directly activates an evaluative process (V). Performance (R^E) relies on both R–O and O–R processes. The necessity of the evaluative pathway in response initiation can be overcome by increasing the contribution of the selection process either by presenting a stimulus separately trained with the outcome (ie, $S^{CS}–O^S$ as in Pavlovian-instrumental transfer experiments) or by strengthening the $O^S–R$ association itself by selective reinforcement through overtraining.

actor such that if the agent moves to a state associated with greater reward (and thus generates a positive prediction error), then the probability of choosing that action in future is increased. Conversely, if the agent moves to a state associated with less reward, this generates a negative prediction error and the probability of choosing that action again is decreased.

Analogies have been drawn between the anatomy and connections of the basal ganglia, and possible neural architectures for implementing reinforcement learning models including the actor/critic (Montague et al, 1996). A key proposal by Montague et al (1996) was that the ventral striatum may have a role as the critic, whereas the dorsal striatum implements the actor. Evidence supporting this conclusion in humans was obtained by O'Doherty et al (2004) who found prediction error signals during performance of an instrumental task but not a Pavlovian task in the dorsal striatum consistent with a role for this region as the actor, while prediction error signals correlated with activity in ventral striatum during both instrumental and Pavlovian conditioning, consistent with the role for this region as the critic. Schonberg et al (2007) subsequently showed that the degree of variance in subjects able to successfully learn an instrumental conditioning task was associated with the extent to which they exhibited prediction errors in dorsal striatum. Furthermore, consistent with the actor/critic proposal, while reward-prediction error activity in ventral striatum was weaker in subjects who failed to learn such a task this ventral striatum prediction error activity did not differ significantly between groups. These results suggest a dorsal/ventral distinction within the striatum, whereas ventral striatum is more concerned with Pavlovian or stimulus-outcome learning, the dorsal stria-

tum is more engaged during the learning of S–R or S–R–outcome associations.

How does a model such as the actor/critic map onto the distinction between goal-directed and habitual reward-predictions described earlier? One possibility, proposed by Daw et al (2005), is that a reinforcement learning model such as the actor/critic is concerned purely with the learning of habitual S–R value signals (see also Balleine et al, 2008). According to this interpretation, action-value signals learned by an actor/critic would not be immediately updated following a change in the value of the reward outcome (such as by devaluation). Instead such an update would occur only after the model re-experiences the reward in its now devalued state and generates prediction errors that would incrementally modulate action values. Daw et al (2005) also proposed an alternative model to the actor/critic to account for the goal-directed component of instrumental learning: a forward model. Unlike reinforcement learning which develops approximate or 'cached' values for particular actions based on earlier experience with those actions, in the 'forward model', values for different actions are worked out on-line by taking into account knowledge about the rewards available in each state, and the transition probabilities between each state and iteratively working out the value of each available option. One property of this model is that value representations should be sensitive to outcome devaluation, because they are computed on-line with respect to the known structure of the decision problem and the incentive value of the outcome; hence, this model accounts for one property of goal-directed learning, namely the sensitivity of goal-directed actions to changes in outcome value. Furthermore, the extent to which the goal-directed 'forward-model' and the habitual RL system controls behavior is argued to depend on an on-going uncertainty-based competition between the two systems so that the system yielding the least uncertain estimates of reward controls behavior at a given point in time, a feature that can be shown to reproduce the differential control of behavior by the goal-directed system early compared with late in training.

Some fMRI evidence in support of the existence of 'forward' model-based value signals in vmPFC has emerged (Hampton et al, 2006). However, although this is a very promising conceptual framework, a number of outstanding issues still need to be resolved. First, although there are several ways of understanding the uncertainty that shifts between goal-directed and habitual control, these appear, nevertheless, to be axiomatically competing processes. Hence, the notion they might at times operate cooperatively is not yet accommodated in this analysis. Second, although it provides an elegant account of the differential sensitivity of under- and overtrained actions to outcome devaluation, it does not, as of yet, account for their differential sensitivity to contingency degradation. Third, a strong implication of this model is that the computational function of dopamine relates specifically to the operation of the habit-like RL system, and not the goal-directed system. Yet, dopaminergic

neurons project heavily to the prefrontal cortex and dorsomedial striatal regions involved in goal-directed behavior as well as to dorsolateral striatum (Gaspar *et al*, 1992; Lidow *et al*, 1991; Lynd-Balta and Haber, 1994). Thus, it seems possible that dopamine has a role in the goal-directed system as well as in the habitual system. Finally, empirical evidence in support of the presence of distinct RL and forward model-based value signals has not yet been forthcoming.

An alternative model of reward-based action selection has been proposed by Frank and Claus (2006). This model provides a literal description of proposed cortico-basal ganglia circuits underlying reward-learning, in which different groups of striatal neurons are proposed to mediate 'go' and 'no-go' responses, learned through the afferent dopaminergic prediction error. In essence this model is a literal implementation of a habit-like reinforcement learning process. The prefrontal cortex is proposed to maintain outcome information during action selection in a working memory, which is used to exert biasing effects on basal ganglia circuitry. The transfer of behavior from goal-directed to habitual control is proposed to take place through the strengthening of habit-like S–R associations in the basal ganglia which overcomes the biasing effects of the prefrontal system. Although this is a promising approach for testing how such dual control processes could be implemented within the architecture of the corticostriatal network, extensions to the current model are required to accommodate all of the known behavioral and neuroanatomical properties of the two systems reviewed earlier, such as the differential sensitivity of these systems to contingency degradation and the differentiation of the striatum into medial (goal-directed) and lateral (habitual) components.

Neural Theories of Action Selection

Recent developments in the neuroanatomy and neuropharmacology of the basal ganglia have spawned a series of theoretical propositions regarding how learning and motivational processes might be integrated within this system. For example, one attractive proposal is that the various ventral and dorsal striatal networks described above, and illustrated in Figure 2a, are hierarchically organized, allowing information to propagate from one level to the next. In particular, the recent description of spiraling connections between the striatum and the midbrain suggest an anatomical organization that can potentially implement interactions between networks (see Figure 2b). As observed by Haber *et al* (2000), striatal neurons send direct inhibitory projections to DA neurons from which they receive reciprocal DA projections, and also project to DA neurons which in turn project to a different striatal area. These latter projections allow feed-forward propagation of information in only one direction, from the limbic networks to associative and sensorimotor networks. On this view, for example, Pavlovian values could influence the effective Pavlovian teaching signal at one level, while enhancing

dopaminergic activity, and hence instrumental performance at the next. This could be revealed in a negative feedback signal from the GABAergic medium spiny projection neurons from the striatum to the DA neurons on the one hand, and the potentiation of the DA signal through disinhibitory projections on the neighboring cortico-basal ganglia network (eg, through GABAergic striatal projection neurons to nigral GABAergic interneurons to DA neurons) (Haber, 2003; Haber *et al*, 2000; Yin *et al*, 2008). This model predicts the progressive involvement of different neural networks during different stages of learning, a suggestion that has received some support (Belin and Everitt, 2008; Everitt and Robbins, 2005; Jueptner *et al*, 1997b; Miyachi *et al*, 2002; Miyachi *et al*, 1997).

Various kinds of evidence support this model; for example, although the feedforward projections from ventral striatum to the DA neurons projecting to ventral striatum (Nauta *et al*, 1978; Nauta, 1989) is relatively greater than the feedback projections to ventral striatum, the opposite is true of the dorsal striatum. As such, the ventral striatal network should be expected to exert greater control over the functions of the dorsal striatum than vice versa. There is in fact some evidence for this prediction; the Pavlovian facilitation of instrumental behavior is greater than the reverse; indeed, evidence suggests that the instrumental actions tend to inhibit, rather than excite, Pavlovian CRs (Ellison and Konorski, 1964; Williams, 1965). As might be expected, therefore, evidence from Pavlovian instrumental transfer experiments is consistent with this model. Thus, for example, dopamine activity appears to be critical for general transfer, which is abolished by DA antagonists and local inactivation of the VTA (Dickinson *et al*, 2000; Murschall and Hauber, 2006), whereas amphetamine-induced sensitization can enhance it (Wyvell and Berridge, 2000). Nevertheless, evidence from specific transfer is more ambiguous; it is partially spared after inactivation of the VTA (Corbit *et al*, 2007), rendered nonselective by inactivation of dorsomedial striatum and abolished by inactivation of the dorsolateral striatum (Corbit and Janak, 2007), suggesting that stimulus control over action selection might be specific to the nigrostriatal projection. Thus, based on these results, the dorsomedial area appears to influence only the specificity of transfer-specific transfer, whereas the DLS could be necessary for any increases in response vigor induced by Pavlovian values.

It is possible to speculate too on the way this model might be thought to account for the influence of outcome values on performance. For example, evidence described above suggests that lesions of accumbens core reduce instrumental performance and also reduce sensitivity to outcome devaluation (Balleine and Killcross, 1994; Corbit *et al*, 2001). Given the role of basolateral amygdala in outcome values and its inputs to accumbens core, there exists at least the possibility that the influence of outcome values on performance is mediated by an amygdala–ventral striatal modulation of the excitatory dopaminergic afferents on dorsomedial striatum from the substantia nigra. Never-

theless, this general confluence of motivational processes through a common dopaminergic mechanism, although parsimonious, does not fit with the growing body of evidence demonstrating that the effects of outcome and Pavlovian values on performance are doubly dissociable (reviewed above). There are, of course, other networks that have been implicated in the way outcome values influence performance. First and foremost among these are the direct projections of basolateral amygdala to medial prefrontal cortex and dorsomedial striatum (Kelley *et al*, 1982). But, of course, there are many other possible routes through which these structures might interact including projections to prefrontal cortex from the ventral striato–pallido–thalamic pathway (Zahm, 2000), and, indeed, directly into dorsal striatum through the thalamo–striatal pathway generally (Groenewegen *et al*, 1999; Haber and McFarland, 2001; Zackheim and Abercrombie, 2005).

Over and above this question of the integration of learning and motivational processes in performance, there is also the question of whether the corticostriatal networks that mediate R–O and S–R learning interact directly. As described above, some evidence suggests that they form a cooperative network. However, there is also evidence of inhibitory or competing interactions between these processes; thus, inactivation of dorsomedial striatum immediately places actions under habitual control (Yin *et al*, 2005), whereas inactivation of the dorsolateral striatum immediately renders actions goal-directed (Yin *et al*, 2006) as if these two processes are always active and merely compete for control of performance (Balleine *et al*, 2009). The generally accepted architecture of the basal ganglia emphasizes the operation of functionally distinct, closed parallel loops connecting prefrontal cortex, dorsal striatum, and the pallidum/substantia nigra, and thalamus that feeds back onto the originating area of prefrontal cortex (Alexander and Crutcher, 1990a; Alexander *et al*, 1986; Nakahara *et al*, 2001). There is, on this view, integration within loops but not vertical integration across loops and, as a consequence, various theories have had to be developed to account for vertical integration; eg, the split loop (Joel and Weiner, 2000) or spiraling midbrain–striatal integration described above (Haber, 2003; Haruno and Kawato, 2006).

In contrast, older theories of striato–pallido–nigral integration proposed that, rather than being discrete, corticostriatal connections converge onto common target regions, particularly in the globus pallidus and substantia nigra, a view that allows naturally for integration between various corticostriatal circuits (Bar-Gad *et al*, 2003; Percheron and Filion, 1991; Yelnik, 2002). Although anatomical studies challenge this view, recent evidence has emerged supporting a hybrid version; that, in addition to the segregated loops, there may also be integration through collateral projections from caudate (or dorsomedial striatum) converging with projections from the putamen (or dorsolateral striatum) onto common target regions in both the internal and external globus pallidus (Nadjar *et al*, 2006; Sadek *et al*, 2007). Whether these converging projections

underlie the integration of the O^S–R and R–O^G associations identified above remains to be established.

However, the processes mediating goal-directed and habitual actions interact, recent data suggest that the basal ganglia are able to maintain these functions in parallel and allow, under some conditions, one or other process either independent control or, under other conditions, both processes to exert cooperative control over choice and decision making. It is important to note that, in suggesting that two distinct learning processes are concurrently engaged, this view implies that the representation of the instrumental outcome has two distinct functions serving both as a reward or outcome, as a part of the action–outcome association underlying goal-directed learning, and also to reinforce an association between the action and antecedent stimuli in habits. How this is achieved is not fully understood, although some evidence suggests that the function of parsing the outcome into both a reward and a reinforcement signal depends on the amygdala (Balleine, 2005, 2009; Balleine and Killcross, 2006; Balleine and Ostlund, 2007d). As described above, considerable evidence has accumulated suggesting that the basolateral amygdala has a central role in encoding the incentive or reward value of the instrumental outcome and, hence, in controlling the performance of goal-directed actions based on the interaction of this evaluative process with the action–outcome association (see section 'The corticolimbic-ventral striatal network and the motivation of decision-making'). Likewise, a number of authors have suggested that the reinforcement signal mediating the acquisition of S–R associations involving the dorsolateral striatum involves the ascending dopaminergic projection arising in the substantia nigra (Faure *et al*, 2005; Reynolds *et al*, 2001), a projection that appears to be at least partly controlled by the central nucleus of the amygdala—see Figure 2b (Gonzales and Chesselet, 1990). Although direct evidence that the CeN has a role in this reinforcement signal (and so in habit formation) has not yet been reported, it is known to be involved in generating general affective responses to rewarding events (Cardinal *et al*, 2002), signals associated with rewarding events (Corbit and Balleine, 2005; Holland and Gallagher, 2003), and in the control of simple S–R associations, such as those involving the performance of orienting responses to stimuli associated with food (El-Amamy and Holland, 2007). By activating both the central and basolateral amygdala, therefore, a single outcome-related event could potentially exert distinct functional effects in the performance of actions in the decision-making situation by controlling the production of independent reward and reinforcement signals that concatenate to distinct regions of striatum to control distinct corticostriatal circuits.

FUTURE DIRECTIONS

Although there are numerous unanswered questions, we have shown here that there is likely to be a great deal of

commonality in the way the cortico-basal ganglia network functions to control adaptive behavior in mammalian species. Thus, we have described here two learning processes and two motivational processes that subserve goal-directed and habitual actions and that not only appear to exert similar behavioral control but also to depend on homologous neural networks in humans and rodents. Of course, we are not proposing that these networks are identical or reducible on a point-to-point basis. It does appear, however, that the functional anatomy and underlying organizational principles of the basal ganglia in humans and rodents are extremely similar and it will be an important task of future research to establish whether and in what ways they differ functionally.

There is, for example, striking similarity in the functional network controlling goal-directed actions. As we describe above, objective and subjective measures of action–outcome learning correlate with activity in regions of medial prefrontal, medial orbital cortex, and the caudate nucleus in humans (Figure 1). Similarly, we reviewed the now considerable evidence that prelimbic cortex and dorsomedial striatum in rats subserves a very similar function. Generally, the homology between the rodent dorsomedial striatum and human caudate nucleus is relatively secure, although there are clearly functional differences in the rostro-caudal plane that remain to be explored (Balleine et al, 2007b). There has, however, been considerably more controversy as to whether the rat possesses similar prefrontal cortical regions to humans and other primates (Preuss, 1995). Nevertheless, this controversy has largely been concerned with narrower issues such as whether there is a rodent homology to primate dorsolateral prefrontal cortex (with Preuss (1995) claiming that there is not and Uylings et al, 2003 arguing that there certainly is) or frontal pole (Wise, 2008), whereas there is growing evidence based on connectivity and density of connections, neurotransmitter types, embryological development, cytoarchitectonic characteristics, and (last but, obviously, not least from, our perspective) functional similarity that rodent prelimbic-medial orbital cortex region is analogous to human ventromedial prefrontal-medial orbital cortex (see Brown and Bowman, 2002; Uylings et al, 2003 for discussion). Generally, however, there has been insufficient systematic research of prefrontal cortex functions in rodent and human subjects using similarly structured tasks. More research along these lines would contribute meaningfully to an understanding of functional homologies in prefrontal cortex.

We also describe commonalities in the network mediating habit learning. Although, given the literature, this is perhaps somewhat less surprising (Graybiel, 2008), it is important to recognize that much of the previous evidence has not only come from very different tasks but also the relevance of these tasks to S–R learning has often largely been inferred from the experimenters' description of the task rather than direct tests of the factors controlling performance. There has, for example, been a large literature linking sensorimotor cortex and its efferents to dorsolateral striatum in rat in various maze learning tasks and other nominally S–R tasks involving discrimination learning. As we describe above, this is, however, something that has now been confirmed more directly using behavioral measures that confirm habitual control of overtrained actions both in rats and, using a very similar task, in humans and found to depend on dorsolateral striatum in rat and on the analogous region of striatum, the putamen, in human. At present, little is known about the structure of habit learning. For example, although reinforcement learning models provide clear predictions on the nature of the reinforcement signal supporting habit learning, it is not currently known if this constitutes the dopamine error signal alone, some integral of that signal involving local neuromodulators of the dopaminergic input to the striatum, or some combination of these or other potential processes. Likewise, although we have argued that amygdala has a central role in both reward and reinforcement processes, we do not know how these twin functions of outcome delivery are parsed neurally.

Although the goal-directed and habit learning processes that support the two major forms of adaptive behavioral control can be observed in both humans and rats and appear to depend on homologous corticostriatal circuits, it is unknown how these forms of learning interact and what conditions modulate whether they cooperate or inhibit/interfere with one another. As presented above, it is clear that under some circumstances they interact; goal-directed and habitual control of performance often appears to be all or none rather than some mixture of the two. In other situations, these processes appear to be temporally related to one another and to function in synergy during the selection, evaluation, and implementation of actions. This will be a critical problem to resolve if we are to formulate accurate models of real life decision making in the course of which ongoing routine actions have often to be suspended, interference between concurrent choices has to be overcome, and new information has to be incorporated into the decision process while, nevertheless, remaining adaptive, at least for the most part.

Finally, we also described considerable evidence suggesting that homologous structures in rat and human mediate the influence of motivational processes on the performance of goal-directed and habitual actions, particularly the influence of Pavlovian cues on choice. Central and lateral OFC and ventral striatum appear to be particularly important for calculating and deploying the influence of Pavlovian values and, as described above, similar structures have also been implicated in rats in this process. This is perhaps most clear in the influence of outcome-specific Pavlovian values on choice in Pavlovian-instrumental transfer; in rodent, this effect depends on a circuit involving accumbens shell, mediodorsal thalamus, and lateral obitofrontal cortex. In humans, to date we have found evidence that the ventral putamen, a region coextensive with the shell, is activated in the presence of cues predicting the same outcome as the action when that action is performed,

and is reduced when an action associated with a different outcome is performed. Similarly, outcome values encoded as a consequence of exposure to, and consummatory contact with, the consequences of actions have been found to depend on amygdala, insular, and medial orbitofrontal cortex in humans and to rely on similar structures in rats. Unfortunately, apart from very few notable exceptions, little of this research has been undertaken using comparable tasks or motivational manipulations organized to influence the same or comparable regulatory systems. Again, it is clearly research along these lines that will make the most rapid progress in understanding functional homologies in motivation. Finally, we do not have a clear understanding of the way that motivational and emotional processes regulate the networks that control choice and decision making in either humans or rodents. The common features of the influence of outcome and Pavlovian values on choice in rodent and human subjects described above points to the fact that similar neural systems organized along similar lines likely underlie this integrative process at least in mammals but, for the moment, it would be pure speculation to attempt to specify those in any greater detail.

ACKNOWLEDGEMENTS

DISCLOSURE

The authors declare that, except for income received from their primary employers, no financial support or compensation has been received from any individual or corporate entity over the past 3 years for research or professional service and there are no personal financial holdings that could be perceived as constituting a potential conflict of interest.

REFERENCES

Abler B, Walter H, Erk S, Kammerer H, Spitzer M (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *Neuroimage* **31**: 790–795.

Adams CD (1981). Variations in the sensitivity of instrumental responding to reinforcer devalaution. *Q J Exp Psychol* **34B**: 77–98. **This paper describes the first experiments demonstrating the differential sensitivity of under- and over-trained actions to outcome devaluation**.

Adams CD, Dickinson A (1981). Instrumental responding following reinforcer devaluation. *Q J Exp Psychol* **33B**: 109–121.

Alexander GE, Crutcher MD (1990a). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci* **13**: 266–271.

Alexander GE, Crutcher MD (1990b). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci* **13**: 266–271.

Alexander GE, DeLong MR, Strick PL (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Ann Rev Neurosci* **9**: 357–381.

Alheid GF (2003). Extended amygdala and basal forebrain. *Ann NY Acad Sci* **985**: 185–205.

Anderson AK, Christoff K, Stappen I, Panitz D, Ghahremani DG, Glover G *et al* (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nat Neurosci* **6**: 196–202.

Antonini A, Moresco RM, Gobbo C, De Notaris R, Panzacchi A, Barone P *et al* (2001). The status of dopamine nerve terminals in Parkinson's disease and essential tremor: a PET study with the tracer [11-C]FE-CIT. *Neurol Sci* **22**: 47–48.

Balleine BW, Delgado MR (2007a). The role of the dorsal striatum in reward and decision-making. *J Neurosci* **27**: 8161–8165.

Balleine BW, Liljeholm M, Ostlund SB (2009). The integrative function of the basal ganglia in instrumental conditioning. *Behav Brain Res* **199**: 43–52.

Balleine BW (1992). Instrumental performance following a shift in primary motivation depends on incentive learning. *J Exp Psychol Anim Behav Process* **18**: 236–250.

Balleine BW (2001). Incentive processes in instrumental conditioning. In: Klein RMS (ed). *Handbook of contemporary learning theories*. LEA: Hillsdale, NJ. pp 307–366.

Balleine BW (2004). Incentive behavior. In: Whishaw IQ, Kolb B (eds). *The Behavior of the Laboratory Rat: A Handbook With Tests*. Oxford University Press: Oxford. pp 436–446.

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

Balleine BW (2009). Taste, disgust and value: taste aversion learning and outcome encoding in instrumental conditioning. In: S. Reilly TRS (ed). *Conditioned Taste Aversion: Behavioral and Neural Processes*. Oxford University Press: Oxford. pp 262–280.

Balleine BW, Dickinson A (1991). Instrumental performance following reinforcer devaluation depends upon incentive learning. *Q J Exp Psychol* **43B**: 279–296.

Balleine BW, Ball J, Dickinson A (1994). Benzodiazepine-induced outcome revaluation and the motivational control of instrumental action in rats. *Behav Neurosci* **108**: 573–589.

Balleine BW, Davies A, Dickinson A (1995). Cholecystokinin attenuates incentive learning in rats. *Behav Neurosci* **109**: 312–319.

Balleine BW, Daw ND, O'Doherty JP (2008). Multiple forms of value learning and the function of dopamine. In: Glimcher PW, Camerer CF, Poldrack RA, Fehr E (eds). *Neuroeconomics: Decision Making and the Brain*. Academic Press: New York.

Balleine BW, Delgado MR, Hikosaka O (2007b). The role of the dorsal striatum in reward and decision-making. *J Neurosci* **27**: 8161–8165.

Balleine BW, Dickinson A (1994). Role of cholecystokinin in the motivational control of instrumental action in rats. *Behav Neurosci* **108**: 590–605.

Balleine BW, Dickinson A (1998a). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* **37**: 407–419. **Describes the development of the cannonical tests of goal-directed action in rodents and the initial evidence for the involvement of prelimbic cortex in the encoding of the action-outcome association**.

Balleine BW, Dickinson A (1998b). The role of incentive learning in instrumental outcome revaluation by specific satiety. *Anim Learn Behav* **26**: 46–59.

Balleine BW, Dickinson A (2000). The effect of lesions of the insular cortex on instrumental conditioning: evidence for a role in incentive memory. *J Neurosci* **20**: 8954–8964.

Balleine BW, Doya K, O'Doherty J, Sakagami M (eds) (2007c). *Reward and Decision Making in Corticobasal Ganglia Networks*. New York Academy of Sciences: New York.

Balleine BW, Killcross AS (1994). Effects of ibotenic acid lesions of the nucleus accumbens on instrumental action. *Behav Brain Res* **65**: 181–193.

Balleine BW, Killcross AS, Dickinson A (2003). The effect of lesions of the basolateral amygdala on instrumental conditioning. *J Neurosci* **23**: 666–675.

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

Balleine BW, Ostlund SB (2007d). Still at the choice point: Action selection and initiation in instrumental conditioning. *Ann NY Acad Sci* **1104**: 147–171.

Bar-Gad I, Morris G, Bergman H (2003). Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Prog Neurobiol* **71**: 439–473.

Barto AG (1992). Reinforcement learning and adaptive critic methods. In: White DA, Sofge DA (eds). *Handbook of intelligent control: Neural, Fuzzy, and Adaptive Approaches*. Van Norstrand Reinhold: New York. pp 469–491.

Barto AG (1995). Adaptive Critics and the Basal Ganglia. In: Houk JC, Davis JL, Beiser BG (eds). *Models of Information Processing in the Basal Ganglia*. MIT Press: Cambridge, MA. pp 215–232.

Baxter MG, Parker A, Lindner CC, Izquierdo AD, Murray EA (2000). Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *J Neurosci* **20**: 4311–4319.

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

Bray S, Rangel A, Shimojo S, Balleine BW, O'Doherty JP (2008). The neural mechanisms underlying the influence of Pavlovian cues on human decision-making. *J Neurosci* **28**: 5861–5866.

Brog JS, Salyapongse A, Deutch AY, Zahm DS (1993). The patterns of afferent innnervation of the core and shell in the 'accumbens' part of the rat ventral striatum:immunohistochemical detection of retrogradely transported fluoro-gold. *J Comp Neurol* **338**: 255–278.

Brown VJ, Bowman EM (2002). Rodent models of prefrontal cortical function. *Trends Neurosci* **25**: 340–343.

Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci Biobehav Rev* **26**: 321–352.

Colwill RC, Rescorla RA (1986). Associative structures in instrumental learning. *Psychol Learn Motiv* **20**: 55–104.

Colwill RM, Rescorla RA (1985). Postconditioning devaluation of a reinforcer affects instrumental responding. *J Exp Psychol: Anim Behav Process* **11**: 120–132.

Colwill RM, Rescorla RA (1990). Effect of reinforcer devaluation on discriminative control of instrumental behavior. *J Exp Psychol Anim Behav Process* **16**: 40–47.

Condé F, Maire-Lepoivre E, Audinat E, F C (1995). Afferent connections of the medial frontal cortex of the rat. II. Cortical and subcortical afferents. *J Comp Neurol* **20**: 567–593.

Corbit LH, Balleine BW (2003a). Instrumental and Pavlovian incentive processes have dissociable effects on components of a heterogeneous instrumental chain. *J Exp Psychol Anim Behav Process* **29**: 99–106.

Corbit LH, Balleine BW (2003b). The role of prelimbic cortex in instrumental conditioning. *Behav Brain Res* **146**: 145–157.

Corbit LH, Balleine BW (2005). Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. *J Neurosci* **25**: 962–970. **Describes initial evidence for two distinct forms of Pavlovian-instrumental transfer: general transfer, mediated by the amygdala central nucleus, and outcome-specific transfer, mediated by the basoloateral amygdala**.

Corbit LH, Janak PH (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 LH, Janak PH, Balleine BW (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.

Corbit LH, Muir JL, Balleine BW (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.

Cromwell HC, Schultz W (2003). Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *J Neurophysiol* **89**: 2823–2838.

Daw ND, Niv Y, Dayan P (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* **8**: 1704–1711. **Proposes a computational account of goal-directed and habitual processes in terms of model-based and model-free reinforcement learning mechanisms**.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006). Cortical substrates for exploratory decisions in humans. *Nature* **441**: 876–879.

Day JJ, Carelli RM (2007). The nucleus accumbens and Pavlovian reward learning. *Neuroscientist* **13**: 148–159.

Day JJ, Wheeler RA, Roitman MF, Carelli RM (2006). Nucleus accumbens neurons encode Pavlovian approach behaviors: evidence from an autoshaping paradigm. *Eur J Neurosci* **23**: 1341–1351.

De Araujo IE, Kringelbach ML, Rolls ET, McGlone F (2003). Human cortical responses to water in the mouth, and the effects of thirst. *J Neurophysiol* **90**: 1865–1876.

de Borchgrave R, Rawlins JN, Dickinson A, Balleine BW (2002). Effects of cytotoxic nucleus accumbens lesions on instrumental conditioning in rats. *Exp Brain Res* **144**: 50–68.

De Wit S, Kosaki Y, Balleine BW, Dickinson A (2006). Dorsomedial prefrontal cortex resolves response conflict in rats. *J Neurosci* **26**: 5224–5229.

Dickinson A (1994). Instrumental conditioning. In: Mackintosh NJ (ed). *Animal Cognition and Learning*. Academic Press: London. pp 4–79. **An important review that evaluates experimental evidence for and against a variety of theories of instrumental conditioning**.

Dickinson A, Balleine BW (1994). Motivational control of goal-directed action. *Anim Learn Behav* **22**: 1–18.

Dickinson A, Balleine BW (2002). The role of learning in the operation of motivational systems. In: Gallistel CR (ed). *Learning, Motivation & Emotion, Volume 3 of Steven's Handbook of Experimental Psychology, Third Edition*. John Wiley & Sons: New York. pp 497–533.

Dickinson A, Balleine BW, Watt A, Gonzales F, Boakes RA (1995). Overtraining and the motivational control of instrumental action. *Anim Learn Behav* **22**: 197–206.

Dickinson A, de Wit S (2003). The interaction between discriminative stimuli and outcomes during instrumental learning. *Q J Exp Psychol B* **56**: 127–139.

Dickinson A, Mulatero CW (1989). Reinforcer specificity of the suppression of instrumental performance on a non-contingent schedule. *Behav Process* **19**: 167–180.

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

Dickinson A, Squire S, Varga Z, Smith JW (1998). Omission learning after instrumental pretraining. *Q J Exp Psychol* **51B**: 271–286. **Describes evidence that, after overtraining, rats are less sensitive to changes in the instrumental contingency, in this case induced by imposing an omission schedule, than when undertrained**.

El-Amamy H, Holland PC (2007). Dissociable effects of disconnecting amygdala central nucleus from the ventral tegmental area or substantia nigra on learned orienting and incentive motivation. *Eur J Neurosci* **25**: 1557–1567.

Elliott R, Newman JL, Longe OA, Deakin JF (2003). Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: a parametric functional magnetic resonance imaging study. *J Neurosci* **23**: 303–307.

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

Everitt BJ, Robbins TW (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat Neurosci* **8**: 1481–1489.

Everitt BJ, Stacey P (1987). Studies of instrumental behavior with sexual reinforcement in male rats (Rattus norvegicus): II. Effects of Preoptic area lesions, castration and testosterone. *J Comp Psychol* **101**: 407–419.

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.

Frank MJ, Claus ED (2006). Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol Rev* **113**: 300–326.

Fudge JL, Emiliano AB (2003). The extended amygdala and the dopamine system: another piece of the dopamine puzzle. *J Neuropsychiatry Clin Neurosci* **15**: 306–316.

Gaspar P, Stepniewska I, Kaas JH (1992). Topography and collateralization of the dopaminergic projections to motor and lateral prefrontal cortex in owl monkeys. *J Comp Neurol* **325**: 1–21.

Gläscher J, Hampton AN, O'Doherty JP (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cereb Cortex* **19**: 483–495.

Gonzales C, Chesselet MF (1990). Amygdalonigral pathway: an anterograde study in the rat with Phaseolus vulgaris leucoagglutinin (PHA-L). *J Comp Neurol* **297**: 182–200.

Gottfried JA, O'Doherty J, Dolan RJ (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J Neurosci* **22**: 10829–10837.

Gottfried JA, O'Doherty J, Dolan RJ (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* **301**: 1104–1107.

Graybiel AM (2008). Habits, rituals, and the evaluative brain. *Annu Rev Neurosci* **31**: 359–387.

Groenewegen HJ, Berendse HW, Wolters JG, Lohman AHM (1990). The anatomical relationship of the prefrontal cortex with the striatopallidal system, the thalamus and the amygdala: Evidence for a parallel organisation. *Prog Brain Res* **85**: 95–118.

Groenewegen HJ, Galis-de Graaf Y, Smeets WJ (1999). Integration and segregation of limbic cortico-striatal loops at the thalamic level: an experimental tracing study in rats. *J Chem Neuroanat* **16**: 167–185.

Guthrie ER (1935). *The Psychology of Learning*. Harpers: New York.

Haber S, McFarland NR (2001). The place of the thalamus in frontal cortical-basal ganglia circuits. *Neuroscientist* **7**: 315–324.

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

Haber SN, Fudge JL, McFarland NR (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J Neurosci* **20**: 2369–2382.

Haber SN, Kim KS, Mailly P, Calzavara R (2006). Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. *J Neurosci* **26**: 8368–8376.

Hall J, Parkinson JA, Connor TM, Dickinson A, Everitt BJ (2001). Involvement of the central nucleus of the amygdala and nucleus accumbens core in

mediating Pavlovian influences on instrumental behaviour. *Eur J Neurosci* **13**: 1984–1992.

Hammond LJ (1980). The effects of contingencies upon appetitive conditioning of free-operant behavior. *J Exp Anal Behav* **34**: 297–304.

Hampton AN, Bossaerts P, O'Doherty JP (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J Neurosci* **26**: 8360–8367.

Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J Neurosci* **28**: 5623–5630.

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

Hatfield T, Han JS, Conley M, Gallagher M, Holland P (1996). Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. *J Neurosci* **16**: 5256–5265.

Hazy TE, Frank MJ, O'Reilly RC (2007). Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philos Trans R Soc Lond B Biol Sci* **362**: 1601–1613.

Hebb DO (1949). *The organization of behavior: a neuropsychological theory.* Wiley: New York.

Hellemans KG, Dickinson A, Everitt BJ (2006). Motivational control of heroin seeking by conditioned stimuli associated with withdrawal and heroin taking by rats. *Behav Neurosci* **120**: 103–114.

Hendersen RW, Graham J (1979). Avoidance of heat by rats: Effects of thermal context on the rapidity of extinction. *Learning Motivation* **10**: 351–363.

Holland PC (2004). Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *J Exp Psychol Anim Behav Process* **30**: 104–117. **Rat study describing the effect of overtraining on both outcome devaluation and Pavlovian-instrumental transfer and demonstrating that, although the influence of outcome value declines, the influence of reward-related stimuli on performance increases**.

Holland PC, Gallagher M (2003). Double dissociation of the effects of lesions of basolateral and central amygdala on conditioned stimulus-potentiated feeding and Pavlovian-instrumental transfer. *Eur J Neurosci* **17**: 1680–1694.

Holland PC, Petrovich GD, Gallagher M (2002). The effects of amygdala lesions on conditioned stimulus-potentiated eating in rats. *Physiol Behav* **76**: 117–129.

Hollerman JR, Schultz W (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat Neurosci* **1**: 304–309.

Hull CL (1943). *Principles of Behavior.* Appleton: New York.

Hutcheson DM, Everitt BJ, Robbins TW, A D (2001). The role of withdrawal in heroin addiction: enhances reward or promotes avoidance? *Nat Neurosci* **4**: 943–947.

Joel D (2001). Open interconnected model of basal ganglia-thalamocortical circuitry and its relevance to the clinical syndrome of Huntington's disease. *Mov Disord* **16**: 407–423.

Joel D, Niv Y, Ruppin E (2002). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Netw* **15**: 535–547.

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 CD, Brooks DJ, Frackowiak RS, Passingham RE (1997a). Anatomy of motor learning. II. Subcortical structures and learning by trial and error. *J Neurophysiol* **77**: 1325–1337.

Jueptner M, Stephan KM, Frith CD, Brooks DJ, Frackowiak RS, Passingham RE (1997b). Anatomy of motor learning. I. Frontal cortex and attention to action. *J Neurophysiol* **77**: 1313–1324.

Kelley AE (2004). Ventral striatal control of appetitive motivation: role in ingestive behavior and reward-related learning. *Neurosci Biobehav Rev* **27**: 765–776.

Kelley AE, Domesick VB, Nauta WJ (1982). The amygdalostriatal projection in the rat–an anatomical study by anterograde and retrograde tracing methods. *Neuroscience* **7**: 615–630.

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

Kim H, Shimojo S, O'Doherty JP (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol* **4**: e233.

Kringelbach ML, O'Doherty J, Rolls ET, Andrews C (2003). Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. *Cereb Cortex* **13**: 1064–1071.

Lauwereyns J, Watanabe K, Coe B, Hikosaka O (2002). A neural correlate of response bias in monkey caudate nucleus. *Nature* **418**: 413–417.

Lehéricy S, Benali H, Van de Moortele PF, Pélégrini-Issac M, Waechter T, Ugurbil K (2005). Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *PNAS* **102**: 12566–12571. **fMRI study showing evidence of a transfer of activity from anterior to posterior dorsal striatum during sequence learning as a function of training**.

Levin BE (1999). Arcuate NPY neurons and energy homeostasis in diet-induced obese and resistant rats. *Am J Physiol* **276**(2 Part 2): R382–R387.

Lidow MS, Goldman-Rakic PS, Gallager DW, Rakic P (1991). Distribution of dopaminergic receptors in the primate cerebral cortex: quantitative autoradiographic analysis using [3H]raclopride, [3H]spiperone and [3H]SCH23390. *Neuroscience* **40**: 657–671.

Lopez M, Balleine BW, Dickinson A (1992). Incentive learning and the motivational control of instrumental performance by thirst. *Anim Learn Behav* **20**: 322–328.

Lynd-Balta E, Haber SN (1994). Primate striatonigral projections: a comparison of the sensorimotor-related striatum and the ventral striatum. **345**: 562–578.

Malkova L, Gaffan D, Murray E (1997). Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rheus monkeys. *J Neurosci* **17**: 6011–6020.

Marsh R, Alexander GM, Packard MG, Zhu H, Wingard JC, Quackenbush G *et al* (2004). Habit learning in Tourette syndrome: a translational neuroscience approach to a developmental psychopathology. *Arch Gen Psychiatry* **61**: 1259–1268.

McClure SM, Berns GS, Montague PR (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron* **38**: 339–346. **fMRI study showing activity in human putamen correlating with prediction errors during Pavlovian conditioning**.

McGeorge AJ, Faull RLM (1989). The organization of the projection from the cerebral cortex to the striatum in the rat. *Neuroscience* **29**: 503–537.

McHaffie JG, Stanford TR, Stein BE, Coizet V, Redgrave P (2005). Subcortical loops through the basal ganglia. *Trends Neurosci* **28**: 401–407.

Miller R (2008). *A theory of the basal ganglia and their disorders.* CRC Press: Boca Raton, FL.

Mink JW (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Prog Neurobiol* **50**: 381–425.

Mirenowicz J, Schultz W (1994). Importance of unpredictability for reward responses in primate dopamine neurons. *J Neurophysiol* **72**: 1024–1027.

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.

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

Mogenson GJ, Jones DL, Yim CY (1980). From motivation to action: functional interface between the limbic system and the motor system. *Prog Neurobiol* **14**: 69–97.

Montague PR, Dayan P, Sejnowski TJ (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* **16**: 1936–1947. **Proposes an analogy between the anatomical structure of the basal ganglia and the pattern of ascending dopaminergic projections, and the actor/critic reinforcement learning model implemented using a temporal difference learning rule**.

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.

Nadjar A, Brotchie JM, Guigoni C, Li Q, Zhou SB, Wang GJ *et al* (2006). Phenotype of striatofugal medium spiny neurons in parkinsonian and dyskinetic nonhuman primates: a call for a reappraisal of the functional organization of the basal ganglia. *J Neurosci* **26**: 8653–8661.

Nakahara H, Doya K, Hikosaka O (2001). Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuomotor sequences—a computational approach. *J Cogn Neurosci* **13**: 626–647.

Nambu A (2008). Seven problems on the basal ganglia. *Curr Opin Neurobiol* **18**: 595–604.

Nauta WJ, Smith GP, Faull RL, Domesick VB (1978). Efferent connections and nigral afferents of the nucleus accumbens septi in the rat. *Neuroscience* **3**: 385–401.

Nauta WJH (1989). Reciprocal links of the corpus striatum with the cerebral cortex and limbic system: A common substrate for movement and thought?. In: Mueller (ed). *Neurology and psychiatry: a meeting of minds*. Karger: Basel. pp 43–63.

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003). Temporal difference models and reward-related learning in the human brain. *Neuron* **38**: 329–337.

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

O'Doherty JP, Kringelbach ML, Rolls ET, Hornak J, Andrews C (2001a). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci* **4**: 95–102.

O'Doherty JP, Rolls ET, Francis S, Bowtell R, McGlone F (2001b). Representation of pleasant and aversive taste in the human brain. *J Neurophysiol* **85**: 1315–1321.

O'Doherty JP, Rolls ET, Francis S, Bowtell R, McGlone F, Kobal G *et al* (2000). Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. *Neuroreport* **11**: 893–897.

Oades RD, Halliday GM (1987). Ventral tegmental (A10) system: neurobiology. 1. Anatomy and connectivity. *Brain Res* **434**: 117–165.

Ongür D, Price JL (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb Cortex* **10**: 206–219.

Ostlund SB, Balleine BW (2005). Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. *J Neurosci* **25**: 7763–7770.

Ostlund SB, Balleine BW (2007a). Instrumental reinstatement depends on sensory- and motivationally-specific features of the instrumental outcome. *Learn Behav* **35**: 43–52.

Ostlund SB, Balleine BW (2007b). Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. *J Neurosci* **27**: 4819–4825.

Ostlund SB, Balleine BW (2008). Differential involvement of the basolateral amygdala and mediodorsal thalamus in instrumental action selection. *J Neurosci* **28**: 4398–4405.

Parkinson JA, Cardinal RN, Everitt BJ (2000). Limbic cortical-ventral striatal systems underlying appetitive conditioning. *Prog Brain Res* **126**: 263–285.

Parkinson JA, Dalley JW, Cardinal RN, Bamford A, Fehnert B, Lachenal G et al (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.

Parkinson JA, Olmstead MC, Burns LH, Robbins TW, Everitt BJ (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 JA, Roberts AC, Everitt BJ, Di Ciano P (2005). Acquisition of instrumental conditioned reinforcement is resistant to the devaluation of the unconditioned stimulus. *Q J Exp Psychol B* **58**: 19–30.

Paton JJ, Belova MA, Morrison SE, Salzman CD (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* **439**: 865–870.

Percheron G, Filion M (1991). Parallel processing in the basal ganglia: up to a point. *Trends Neurosci* **14**: 55–59.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442**: 1042–1045. **Pharmacological fMRI study showing that modulation of dopamine even systemically in humans directly influences prediction error signals in human striatum during instrumental learning, providing evidence for a dopaminergic involvement in the BOLD prediction error signal.**

Petrovich GD, Setlow B, Holland PC, Gallagher M (2002). Amygdalo-hypothalamic circuit allows learned cues to override satiety and promote eating. *J Neurosci* **22**: 8748–8753.

Pickens CL, Saddoris MP, Setlow B, Gallagher M, Holland PC, Schoenbaum G (2003). Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *J Neurosci* **23**: 11078–11084.

Plassmann H, O'Doherty J, Rangel A (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J Neurosci* **27**: 9984–9988. **fMRI study showing that medial orbitofrontal cortex (part of vmPFC) correlates with decision values for food rewards, a finding consistent with a role for this region in encoding the incentive value of goal-states.**

Preuss TM (1995). Do rats have prefrontal cortex? The Rose-Woolsey-Akert program reconsidered. *J Cogn Neurosci* **7**: 1–24.

Reardon F, Mitrofanis J (2000). Organisation of the amygdalo-thalamic pathways in rats. *Anat Embryol (Berl)* **201**: 75–84.

Rescorla RA (1994). Transfer of instrumental control mediated by a devalued outcome. *Anim Learn Behav* **22**: 27–33. **Reports the important behavioral demonstration in rodents that devaluation of the outcome associated with a stimulus does not affect the ability of that stimulus to select actions that are associated with that outcome.**

Rescorla RA, Wagner AR (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In: Black AH, Prokasy WF (eds). *In: Classical conditioning II: Current research and theory*. Appleton-Century-Crofts: New York. pp 64–99.

Reynolds JN, Hyland BI, Wickens JR (2001). A cellular mechanism of reward-related learning. *Nature* **413**: 67–70. **Describes evidence that the substantia nigra dopaminergic afferents on dorsolateral striatum subserve the reinforcement signal supporting sensory-motor association.**

Rodgers KM, Benison AM, Klein A, Barth DS (2008). Auditory, somatosensory, and multisensory insular cortex in the rat. *Cereb Cortex* **18**: 2941–2951.

Rolls ET, Kringelbach ML, de Araujo IE (2003). Different representations of pleasant and unpleasant odours in the human brain. *Eur J Neurosci* **18**: 695–703.

Sadek AR, P.J. M, Bolam JP (2007). A single-cell analysis of intrinsic connectivity in the rat globus pallidus. *J Neurosci* **27**: 6352–6362.

Saint-Cyr JA, Taylor AE, Nicholson K (1995). Behavior and the basal ganglia. *Adv Neurol* **65**: 1–28.

Schoenbaum G, Chiba AA, Gallagher M (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat Neurosci* **1**: 155–159.

Schonberg T, Daw ND, Joel D, O'Doherty JP (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J Neurosci* **27**: 12860–12867.

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

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

Seeley RJ, Matson CA, Chavez M, Woods SC, Dallman MF, Schwartz MW (1996). Behavioral, endocrine, and hypothalamic responses to involuntary overfeeding. *Am J Physiol* **271**(3 Part 2): R819–R823.

Shanks DR, Dickinson A (1991). Instrumental judgment and performance under variations in action-outcome contingency and contiguity. *Mem Cognition* **19**: 353–360.

Shidara M, Aigner TG, Richmond BJ (1998). Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J Neurosci* **18**: 2613–2625.

Small DM, Gregory MD, Mak YE, Gitelman D, Mesulam MM, Parrish T et al (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* **39**: 701–711.

Sutton RS, Barto AG (1998). *Reinforcement Learning: An Introduction*. MIT Press: Cambridge, MA.

Talmi D, Seymour B, Dayan P, Dolan RJ (2008). Human pavlovian-instrumental transfer. *J Neurosci* **28**: 360–368. **Human fMRI study showing that general Pavlovian-instrumental transfer effects are associated with activity in human nucleus accumbens, consistent with effects found in the rodent.**

Tanaka SC, Balleine BW, O'Doherty JP (2008). Calculating consequences: Brain systems that encode the causal effects of actions. *J Neurosci* **28**: 6750–6755.

Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat Neurosci* **7**: 887–893.

Tanaka SC, Samejima K, Okada G, Ueda K, Okamoto Y, Yamawaki S et al (2006). Brain mechanism of reward prediction under predictable and unpredictable environmental dynamics. *Neural Netw* **19**: 1233–1241.

Tremblay L, Schultz W (1999). Relative reward preference in primate orbitofrontal cortex. *Nature* **398**: 704–708.

Tricomi E, Balleine BW, O'Doherty JP (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur J Neurosci* **29**: 2225–2232. **Provides behavioral evidence that in humans as in rodents, over-training can render the control of actions habitual, as assessed by sensitivity to outcome devaluation. Morevoer activity in human posterior dorsal striatum was found to track the behavioral development of habits.**

Uylings HB, Groenewegen HJ, Kolb B (2003). Do rats have a prefrontal cortex? *Behav Brain Res* **146**: 3–17.

Valentin VV, Dickinson A, O'Doherty JP (2007). Determining the neural substrates of goal-directed learning in the human brain. *J Neurosci* **27**: 4019–4026. **fMRI study using an outcome devaluation manipulation showing that human ventromedial prefrontal cortex exhibits a response profile consistent with goal-directed learning.**

Wang R, Liu X, Hentges ST, Dunn-Meynell AA, Levin BE, Wang W et al (2004). The regulation of glucose-excited neurons in the hypothalamic arcuate nucleus by glucose and feeding-relevant peptides. *Diabetes* **53**: 1959–1965.

Wang SH, Ostlund SB, Nader K, Balleine BW (2005). Consolidation and reconsolidation of incentive learning in the amygdala. *J Neurosci* **25**: 830–835.

Wasserman EA, Chatlosh DL, Neunaber DJ (1983). Perception of causal relations in humans: Factors affecting judgments of response-outcome contingencies under free-operant procedures. *Learn Motivation* **14**: 406–432.

Wassum K, Ostlund S, Maidment N, Balleine B (2009). Distinct opioid circuits determine the palatability and desirability of rewarding events. *PNAS* **106**: 12512–12517.

Williams BA (1989). The effect of response contingency and reinforcement identity on response suppression by alternative reinforcement. *Learn Motivation* **20**: 204–224.

Williams DR (1965). Classical conditioning and incentive motivation. In: Prokasy WF (ed). *Classical Conditioning*. Appleton-Century-Crofts: New York. pp 340–357.

Winston JS, Gottfried JA, Kilner JM, Dolan RJ (2005). Integrated neural representations of odor intensity and affective valence in human amygdala. *J Neurosci* **25**: 8903–8907. **fMRI study providing evidence that human amygdala activation reflects an interaction between valence and intensity of an olfactory stimulus rather than merely one or other of those alone,**

suggestive of a role for this region in processing affective value for both positive and negative stimuli.

Wise SP (2008). Forward frontal fields: phylogeny and fundamental function. *Trends Neurosci* **31**: 599–608.

Woods SC, Schwartz MW, Baskin DG, Seeley RJ (2000). Food intake and the regulation of body weight. *Annu Rev Psychol* **51**: 255–277.

Woodson JC, Balleine BW (2002). An assessment of factors contributing to instrumental performance for sexual reward in the rat. *Q J Exp Psychol B* **55**: 75–88.

Wyvell CL, Berridge KC (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.

Yelnik J (2002). Functional anatomy of the basal ganglia. *Mov Disord* **17**(Suppl 3): S15–S21.

Yin HH, Ostlund SB, Balleine BW (2008). Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. *Eur J Neurosci* **28**: 1437–1448.

Yin HH, Knowlton BJ, Balleine BW (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci* **19**: 181–189. **Describes direct evidence for the involvement of dorsolaterla striatum in habit learning using outcome devaluation as a test for habitual control**.

Yin HH, Knowlton BJ, Balleine BW (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav Brain Res* **166**: 189–196.

Yin HH, Ostlund SB, Knowlton BJ, Balleine BW (2005). The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci* **22**: 513–523. **Describes initial evidence for the involvement of the dorsomedial striatum in goal-directed learning in rats**.

Zackheim J, Abercrombie ED (2005). Thalamic regulation of striatal acetylcholine efflux is both direct and indirect and qualitatively altered in the dopamine-depleted striatum. *Neuroscience* **131**: 423–436.

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