Recent research in instrumental conditioning has focused on the striatum, particularly the role of the dorsal striatum in the learning processes that contribute to instrumental performance in rats. This research has found evidence of what appear to be parallel, functionally and anatomically distinct circuits involving dorsomedial striatum (DMS) and dorsolateral striatum (DLS) that contribute to two independent instrumental learning processes. Evidence suggests that the formation of the critical action–outcome associations mediates goal-directed action localization to the dorsomedial striatum, whereas the sensorimotor connections that control the performance of habitual actions are localized to the dorsolateral striatum. In addition to the dorsal striatum, these learning processes appear to engage distinct cortico-striatal networks and be embedded in a complex of converging and partially segregated loops that constitute the cortico-striatal thalamo-cortical feedback circuit. As the entry point for the basal ganglia, cortical circuits involving the dorsal striatum are clearly in a position to control a variety of motor functions but, as recent studies of various neurodegenerative disorders have made clear, they are also involved in a number of cognitive and executive functions including action selection, planning, and decision-making.

In what follows we will briefly review the behavioral and neural evidence for these claims before considering two important issues involving, first, evidence for the distinct sources of these influences on performance at both a cortical and a striatal level and, second, the evidence demonstrating their integration in implementing a course of action and the role of the basal ganglia in this process. It will be noted that we plan here to focus primarily on the processes contributing specifically to instrumental conditioning. Other recent reviews have discussed the relationship between instrumental and Pavlovian conditioning processes and their neural bases in more detail and the interested reader is referred there for further discussion (cf. [7,14–16]).
1. Goal-directed action

1.1. Cognition, behavioral control and Pavlovian conditioning

Paradoxically, although the cognitive control of behavior has been of increasing interest to neuroscientists, research in this area has focused predominantly on predictive learning in Pavlovian conditioning paradigms such as fear conditioning and eye-blink conditioning. There is, however, no necessary relationship between cognition and the performance of the Pavlovian conditioned response (CR). Indeed, although it is not generally recognized, at an adaptive level a cognitive mechanism is of little functional value to a purely Pavlovian animal because the production of the CR is under the control of the CS–US association and is demonstrably not determined by a direct relationship between the CR and the US (e.g. [17–19]). As a consequence, although the production of the CR is clearly influenced by the nature of the CS–US association, no amount of refinement in the cognitive representation of the CS, US or their relationship can increase the ability of an animal to control the direction of the CR. In fact, a cognitive mechanism can only exert a functional effect on behavior when coupled to a process capable of modifying, withholding or reversing the direction of actions on the basis of that information, something that demands greater behavioral control than the system mediating Pavlovian conditioning provides (cf. [20,21] for further discussion).

For similar reasons, the Pavlovian paradigm can provide only a limited animal model of the effects of neuropathology on, so-called, executive functions in humans and that evidence suggests depend upon the integrity of various prefrontal-subcortical circuits [22–24]. Deficits in executive function have been generally described as comprising multiple components usually including volition, planning, and purposive action [25], capacities that fall outside the Pavlovian domain. The limbic cortices appear to be particularly heavily affected in executive dysfunction and several investigators have proposed that distinct constellations of symptoms may reflect the disconnection of this cortical area from specific subcortical regions such as the mediodorsal thalamus (in Alzheimers, e.g. [26]), areas of the striatum (in Parkinsons, Huntington and obsessive compulsive disorders, e.g. [27–31]), and the amygdala (in various emotional disorders, e.g. [32]). A disturbing feature highlighted in recent work is the increasing evidence for the early onset of many of the dysfunctions associated with these disorders, something that suggested to Brown and Marsden [27], amongst others, that even quite substantial motor deficits involving tremor and choreic symptoms may partially reflect a disorder in the sustained functioning of a prefrontal-basal ganglia-cortical feedback network engaged during planning, response selection and initiation. However, studying normal and pathological executive functions will require models of behavioral control that go beyond predictive learning to capture the processes engaged during the acquisition and implementation of new behavioral strategies.

1.2. Cognitive and motivational control of actions

Given these limits of Pavlovian processes, it is important to note that instrumental conditioning in rodents has been found to provide an alternative and quite accurate model of executive control generally and of human goal-directed action in particular. Models of human action (e.g. [33–36]) have tended to focus on two critical determinants of goal-directed actions: (1) their dependency on the experienced causal relation between acting (or not acting) and the occurrence of some consequence; and (2) the sensitivity of these actions to changes in the desirability of the consequences or goal of an action. From this perspective, actions that persist even when causally unrelated to their consequences or when those consequences are demonstrably no longer valued should not be regarded as goal-directed.

As we pointed out some time ago [5], this “desire plus belief” characterization of human actions can be used to distill two criteria, what we have called the contingency and the goal criteria, for the detection of goal-directed actions in any species. Since that time we have accumulated considerable evidence suggesting that, for the most part, the performance of instrumental actions by rodents satisfies these criteria. Not only are these instrumental actions highly sensitive to changes in the value of their associated outcome, i.e. post-training devaluation often produces profound changes in the subsequent rate of performance (cf. [6,7,10] for reviews), there is also considerable evidence suggesting that, unlike Pavlovian CR’s, these actions are sensitive to changes in the causal relation to their consequences; generally, rats will stop responding if performance no longer delivers the instrumental outcome and will stop responding even faster if their responding now cancels an otherwise freely available food [37,38]. Furthermore, using a schedule developed by Hammond [39], in which the probability of an outcome given a response (i.e. p(O/R)) and the probability of an outcome in the absence of that response (p(O/noR)) can be independently manipulated, we, amongst others, have reported clear evidence that performance declines as the latter probability increases, even when action–outcome contiguity (i.e. p(R/O)) is kept constant and at a rate that ordinarily maintains substantial levels of performance [8,9,40–43].

Given the clear sensitivity of actions to changes in the probability of outcome delivery it might also be expected that performance would also be sensitive to information concerning the likelihood of earning a particular outcome. And, indeed, there is considerable evidence that stimuli associated with rewarding events can exert quite specific effects on outcome selection and on choice in studies assessing Pavlovian-instrumental transfer. What has also become clear, however, is that this effect does not depend on the mere association of the CS and US but on the information that the stimulus provides about the forthcoming US. Delamater [44], for example, has demonstrated that, although reducing the predictive validity of a cue with respect to the specific US with which it was associated had only a mild effect on the performance of the Pavlovian CR, it completely abolished the influence of that cue on instrumental choice performance in a test of Pavlovian-instrumental transfer. These data suggest that the influence of reward-related stimuli on choice between actions is based on the information that these stimuli convey.

Further evidence along these lines comes from the recent finding that the processes by which reward itself and stimuli that predict reward control goal-directed action differ. Importantly, many theories of goal-directed action either do not distinguish between these processes or view one as ancillary to the other [45–49]. For example, in reinforcement learning the value of the instrumental outcome is coextensive with the value of stimuli or states that predict that outcome and changes in performance following changes in value are, therefore, determined by this common evaluative process [49–51]. This is also true of theories derived from economics, such as utility theory, where the expected value of an action is a product of the amount and the probability of reward or, for expected utility, a weighted average calculated from the utility in each state. In fact, contrary to these suggestions, the influence of the experienced outcome value (as assessed by outcome revaluation) and of expected value based on cues that predict reward (assessed using Pavlovian-instrumental transfer) on the performance of goal-directed actions has been doubly dissociated both neurally [52] and behaviorally [53,54].
It is interesting to note that establishing the distinct neural circuits that mediate the effects of changes in reward and of reward prediction may provide a basis for characterizing the distinct influences of motivational/emotional processes and of cognition on decision-making (see particularly [55]). Whereas changes in reward value induced by motivational manipulations are based on changes in the emotional response associated with the rewarding event [10,11,56], the information conveyed by cues associated with reward, particularly with respect to the relative validity of their predictions, may reflect the role that information regarding the consequences of acting (such as that conveyed by advertising) can have on action selection. Certainly, changes in outcome value appear to have very little effect on the biasing effects that reward-related cues have on choice [15,54,57]. Furthermore, consistent with the claim that it is the information that cues provide about outcomes rather than their ability to act as surrogates for outcome evaluation that determine their influence on choice, several of the neural structures found to play a role in outcome specific Pavlovian-instrumental transfer effects, notably the basolateral amygdala, mediodorsal thalamus and orbitofrontal cortex, have also been found to mediate the predictive validity of the Pavlovian CS, i.e. lesions of these structures each render rats insensitive to degradation of specific Pavlovian CS–US contingencies [58,59].

1.3. Neural bases of goal-directed action

Psychologically speaking, the learning and memory processes underlying goal-directed actions should clearly be regarded as declarative; choice performance clearly reflects the ability of animals to express their knowledge of various action–outcome relations in the face of changing expectations of reward [60]. Nevertheless, despite arguments regarding the function of the hippocampus in declarative learning of this kind [61,62,139], in several series of studies we were unable to find any clear evidence for the involvement of the hippocampus or its projections through anterior thalamus in instrumental learning [43,63,64]. These early experiments did, however, find evidence for the involvement of the mediodorsal thalamus as well as one of its main cortical efferents – the prelimbic region of the medial prefrontal cortex (PL) – in this form of learning. Unlike the hippocampus, cell body lesions of these areas were effective in abolishing rats’ sensitivity to both outcome devaluation and to selective degradation of the instrumental, action–outcome contingency [89,53,64]. A summary of experiments assessing the influence of pretraining lesions of various afferents and efferents of the prefrontal cortex on these tests is presented in Table 1.

Recently we have found evidence that the involvement of the prefrontal cortex in goal-directed learning is phase-limited (see also [65,66]). In a recent series we found clear evidence that only damage to the PL made prior to instrumental training had any effect on conditioning; lesions made after training was complete had no effect on outcome devaluation [67]. This suggested to us that, although the PL is clearly involved in goal-directed learning it is not the locus of the action–outcome association. The PL has two well documented striatal efferents; one arising predominantly in layer II and projecting to the core of the nucleus accumbens [68] and a second arising predominantly in layers V/VI and projecting to the dorsomedial or associative striatum (DMS) [69]. The results of other work had led us to believe that the former plays an important role in instrumental performance but not in instrumental learning [52]. As such we turned our attention to the other projection to the dorsomedial striatum.

In fact, the DMS is an excellent candidate for the locus of the plasticity mediating the encoding of the action–outcome association in instrumental conditioning. As illustrated in Fig. 1, it is a critical component in the associative cortico-basal ganglia circuit and receives inputs from association cortices such as the PL as well as the premotor or medial agranular cortex involved in the action monitoring and programming implicated in executive processes [70,71] and projections from the DMS are in a position to influence downstream motor control networks in the brainstem as well as the motor thalamo-cortical reentrant network [70]. The posterior part of the DMS also receives inputs from the basolateral amygdala [72], a structure that, according to recent evidence, mediates the assignment of incentive value to the consequences of instrumental actions [73,74]. In accord with this suggestion, electrophysiological studies measuring neural activity in the associative striatum or caudate nucleus in primates, the homologue of the DMS in rats, have reported that neural activity in this region correlated with the performance of motor movements and can be modulated by the expectancy of reward [75,76]. More directly, in a recent series of experiments we found direct evidence that, in contrast to manipulations of prefrontal cortex, both pre- and post-training cell body lesions of the DMS as well as local inactivation of this area induced by infusions of the GABA-A agonist muscimol, reduced the sensitivity of rats’ instrumental performance both to shifts in the action–outcome contingency and to post-training outcome devaluation [77].

The suggestion that the DMS is the locus of action–outcome encoding in instrumental conditioning contrasts with other recent claims that the ventral [78] or the posterolateral striatum [79] mediates learning critical to the acquisition of goal-directed actions. Nevertheless, these studies only assessed changes in instrumental performance and did not directly assess changes in the content of learning. Furthermore, as presented in Table 1, we have found that lesions of the ventral striatum, although sometimes effective in influencing performance, do not affect the rat’s sensitivity to changes in the action–outcome contingency. In a second recent series, however, we used well-established behavioral assays that unambiguously distinguish action–outcome learning from other types of learning to assess the role of the DMS in the formation of action–outcome associations [80]. Given the evidence that NMDA receptor (NMDAR) activation is involved in long-term plasticity such as long-term potentiation in the dorsal striatum [81,135], we proposed that action–outcome encoding requires activation of NMDARs in the DMS. This hypothesis was tested in rats that, after a period of pre-training, were given a bilateral infusion of

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Summary of the effects of excitotoxic lesions of various components of the corticostriatal network on sensitivity to selective contingency degradation and outcome devaluation in instrumental conditioning.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>Contingency degradation</td>
</tr>
<tr>
<td>PL</td>
<td>×</td>
</tr>
<tr>
<td>OFC</td>
<td>×</td>
</tr>
<tr>
<td>DMS</td>
<td>×</td>
</tr>
<tr>
<td>DLS</td>
<td>√</td>
</tr>
<tr>
<td>MDT</td>
<td>×</td>
</tr>
<tr>
<td>ANT</td>
<td>√</td>
</tr>
<tr>
<td>NACsh</td>
<td>√</td>
</tr>
<tr>
<td>NACco</td>
<td>√</td>
</tr>
<tr>
<td>HPC</td>
<td>√</td>
</tr>
<tr>
<td>EC</td>
<td>×</td>
</tr>
<tr>
<td>ACC</td>
<td>?</td>
</tr>
<tr>
<td>GI</td>
<td>√</td>
</tr>
</tbody>
</table>

Abbreviations: √: normal ×: deficit; PL: prelimbic area; OFC: orbitofrontal cortex; DMS: dorsomedial striatum; DLS: dorsolateral striatum; MDT: mediodorsal thalamus; ANT: anterior thalamic nuclei; NACsh: nucleus accumbens shell; HPC: hippocampal formation; EC: entorhinal cortex; ACC: anterior cingulated cortex; GI: gustatory insular cortex.
either a selective NMDAR antagonist (APV), or vehicle prior to a single learning session in which they were trained to press two levers for distinct food outcomes. The next day the rats were tested using an outcome devaluation protocol, i.e. they were allowed to consume one of the two outcomes for 1 hr before a choice extinction test was given on the two levers. We found, first that APV immediately prior to training did not affect performance either during training or test  

either a selective NMDAR antagonist (APV), or vehicle prior to a single learning session in which they were trained to press two levers for distinct food outcomes. The next day the rats were tested using an outcome devaluation protocol, i.e. they were allowed to consume one of the two outcomes for 1 hr before a choice extinction test was given on the two levers. We found, first that APV immediately prior to training did not affect performance either during training or testing but strongly attenuated the ability of the rats to use changes in outcome value to modify their instrumental performance, i.e. they appeared not to have encoded the specific action–outcome associations to which they were exposed during training. Furthermore, in subsequent experiments we found both that APV infused immediately after training did not have this effect on action–outcome encoding, nor did the infusion of APV into adjacent dorsolateral striatum (DLS) [80].

1.4. Causal learning and the cortico-striatal network

One element of the rodent data that has remained open to critique has been the difficulty convincingly to demonstrate the active involvement of this prefrontal cortical–dorsomedial striatal network in the translation of causal knowledge about actions and their consequences into goal-directed learning. In a recent study, however, we tried to address this directly not in rats but in human subjects [82]. We trained humans on a free-operant button-pressing task in which they could earn money as the outcome and from these sessions we extracted both an objective measure of the instrumental contingency (i.e. the rate of button pressing and of outcome delivery through time) and a subjective measure that we took by asking the subjects to rate on a 100 point scale how causal they thought their actions were. We gave the subjects a number of sessions of training and compared the variation in these measures of the objective and subjective contingency as well as changes in the activity of neural structures by scanning subjects using functional magnetic resonance imaging while they pressed a button and earned money as this response–reward relationship changed over time.

We found that the subjects’ judgments about the causal efficacy of their actions varied positively and significantly with the objective contingency between the rate of button pressing and the amount of money they earned. Furthermore, as we reported in the rodent, neural responses in medial frontal cortex and dorsomedial striatum were modulated as a function of contingency and these regions altered their activity during periods when actions were highly causal compared with when they were not. Moreover, we found that the medial prefrontal cortex tracked local changes in the correlation of action and outcome rates, implicating this region in the on-line computation of contingency [82]. Hence, just as we have found in rodents, in this study a network involving the medial prefrontal and medial orbital cortices together with the dorsomedial striatal was implicated in detecting the causal relationship between actions and their consequences consistent with the claim that these structures act together to during the acquisition of goal-directed actions to encode specific action–outcome associations.

2. Habitual action

Input to the striatum from sensorimotor cortex, particularly primary motor and somatosensory cortices (cf. [83,84]), appears to be involved in what is commonly thought to be a completely different form of performance process involving the control of actions by antecedent stimuli through a traditional S–R/reinforcement associative mechanism. In recent years this process has been argued to form the core of a distinct functional capacity involving the formation of habitual actions. Habits are revealed particularly in their persistence, even in the face of sometimes quite extreme negative consequences, and in their sensitivity to the motivational functions of reward-related cues [85]. Many of the ideas that have been expressed in recent papers, particularly those linking habit learning with various addictive behaviors (e.g. [86–89,142]), have their root in now classical theories of habit learning, associated most notably with Hull [2], that explain the acquisition of actions instrumental to gaining access to rewarding events in terms of the operation of a S–R/reinforcement architecture.

From this perspective, rewarding events reinforce or strengthen associations between continguously active sensory and motor processes allowing the sensory process to elicit the motor response in a manner that is no longer regulated by its consequences. As such, the determining features of habit learning can be readily distinguished from those of goal-directed action: (1) the lack of regulation by the consequences of an action suggests that habitual
actions should be insensitive to post-training outcome devaluation; (2) the emphasis on stimulus-response contiguity suggests that, rather than reflecting the operation of an error correction learning rule, a particular S–R association is strengthened whenever the response is reinforced in the stimulus, irrespective of the specific outcome delivered or other stimuli present.

In line with these proposed features of habits, there is considerable evidence that when either overtrained [90] or trained on interval schedules of reinforcement [91,92]—i.e. whenever changes in the rate of reward are constrained—instrumental performance becomes insensitive to changes in outcome value. Likewise, evidence suggests that, unlike goal-directed actions, habitual actions are relatively insensitive to changes in the action–outcome contingency. For example, Dickinson et al. [38] found that, when asked to withhold the performance of a previously reinforced lever press action in order to get access to sugar, rats that had been undertrained were able to do so whereas those that had been overtrained were not. This finding has been replicated both in a different kind of training situation [93] and in a different species (i.e. mice; [94]) and suggests, in line with the features described above, that habitual actions are relatively insensitive to changes in the relationship between action and outcome.

Interestingly, evidence suggests that a cortico-striatal network parallel to that implicated in goal-directed action involving somatomotor cortices together with the dorsolateral striatum in rodents may mediate the transition to habitual decision processes associated with S–R learning (see Fig. 1; [95,96]). Changes in the DLS appear to be related to learning-related processes [97–99] and to be coupled to changes in plasticity as behavioral processes become less flexible [99,100]. Correspondingly, whereas overtraining causes performance to become insensitive to outcome devaluation, we have found that lesions of DLS reverse this effect, rendering performance to become insensitive to outcome devaluation, we have found that lesions of DLS reverse this effect, performing rendering performance once again sensitive to devaluation treatments [101]. Likewise, we have found that muscimiol inactivation of DLS renders otherwise habitual performance sensitive to changes in the action–outcome contingency [80].

It is worth noting, however, that, because these are essentially tests for features of habits that distinguish them from goal-directed actions, they establish evidence for habits by default and, while the strength of the S–R association likely correlates strongly with the degree of insensitivity to devaluation and contingency manipulations, insensitivity to these manipulations does not directly assess S–R learning per se. However, positive evidence for the involvement of the DLS in S–R learning has been reported using instrumental conditioning procedures that encourage animals to form S–R relationships in order to solve complex discrimination problems. For example, Featherstone and McDonald [102,103] have shown that lesions of the DLS impair both the acquisition and performance of a simple discrimination task in which lever presses are reinforced during presentations of one stimulus (S+) but not during presentations of another stimulus (S−). Consistent with both an S–R structure and with the putative role of the DLS in this associative process, a salient feature of stimulus-controlled instrumental performance is the general failure to respond that can be generated by DLS lesions (see, for example, [103]) in contrast to their effects on simple free-operant tasks (cf. [101]). Indeed, studies assessing the role of the striatum in basic motor behaviors have shown that lateral striatal lesions can cause severe impairments in both the initiation and amplitude of movements such as forelimb reaches (e.g. [138]). However, these impairments of basic movement can make the results of experiments aimed at investigating specific learning deficits difficult to interpret. For example, in Featherstone and McDonald [103] assessment of the influence of post-training DLS lesions on a simple discrimination, DLS lesioned animals responded significantly less than sham controls on S+ trials, and did not differ from the sham group in their responding on S− trials. Although consistent with a deficit in stimulus-controlled actions, it is also possible, because post-training responding during the S− was close to zero in both groups, that the lack of a difference during the S− reflected a floor-effect.

In an attempt better to distinguish failures to respond from failures of stimulus control over actions during this kind of discrimination task, we trained animals on a conditional discrimination in which the two discriminative stimuli supported equal levels of responding but on different levers. We then assessed the dose-dependent influence of muscimol-induced inactivation of the DLS on response initiation and discriminatory accuracy (correct responses/total responses). During training, rats were required to press the right lever (R1) in response to one auditory cue (S1) and the left lever (R2) in response to a second auditory cue (S2). All correct lever presses were reinforced with a grain pellet outcome (O1), and trials were terminated (i.e. levers retracted) by the first response, whether it was correct or incorrect, and training continued until all animals had reached at least 70% accuracy. As such the structure of the task constituted a bidirectional discrimination problem, viz:

S1 : R1–O1, R2−; S2 : R1−, R2–O1

Although it is possible that animals form hierarchical S:R–O associations to solve this task, the simplest solution involves forming two simple S–R associations: S1–R1 and S2–R2. Of course, to the extent that rats utilize conditional R–O associations we should not anticipate effects of DLS inactivation on this task. If, however, the rats acquired and utilized the simpler S–R solution to this problem then we should find evidence specifically of a failure of discrimination after DLS inactivation resulting from an inactivation-induced inability to use the S–R associations encoded during training.

To assess this prediction we conducted three separate tests in each of which the Long–Evans rats that we used as subjects (n = 10) were required to perform the discrimination task in each of three conditions: after bilateral infusion into the DLS of a high dose of muscimol (0.5 μg, 0.5 μl per hemisphere), a low dose of muscimol (0.25 μg, 0.5 μl per hemisphere) or of vehicle (0.5 μl per hemisphere), in counterbalanced order. Fig. 2 shows the discrimination accuracy and initiation failures in the upper and lower panels, respectively, for the three test conditions. ANOVA found an effect of drug dose on both discrimination, F(2, 18) = 23.0, and response initiation, F(2, 18) = 27.7. Nevertheless, as is clear from Fig. 2, these effects altered across the course of testing. In the high dose condition a clear and immediate loss of discrimination accuracy was observed. Although this was difficult to dissociate from the performance effects of the muscimol infusion on response initiation, it should be noted that, for the first 20 trials or so, the rats were responding on more than half the trials and yet showed no evidence of accurate discrimination. This pattern was even clearer at the lower dose where we found that discrimination accuracy was reduced significantly by the second block of trials (p < 0.05) at a point when response initiation did not differ from vehicle controls (p > 0.05). From this point discrimination at the low dose fell essentially to chance levels with only very minimal effects on response initiation; by the third or fourth block of 10 trials during the test, the rats were completing approximately 75% of the trials and yet their discrimination accuracy had fallen to chance and did not differ from accuracy in the high dose group (p > 0.05). Indeed, both groups differed from the vehicle controls from the second block onwards (all p’s < 0.05).

These results allow us to more conclusively separate discrimination failures from more general initiation failures, and provide
specifically, a circuit mediating goal-directed learning and involving the dorsal striatum described above have identified two distinct areas. 

3. The relationship of goal-directed and habit processes

3.1. Competition or cooperation?

Together, the findings from the experimental investigations of the dorsal striatum described above have identified two distinct functional systems within adjacent regions of dorsal striatum: specifically, a circuit mediating goal-directed learning and involving the dorsomedial striatum and a circuit mediating habit or procedural learning and involving the dorsolateral striatum. Furthermore, at least at the level of the striatum these functions appear to be independent; damage to dorsolateral but not dorsomedial striatum renders otherwise habitual actions goal-directed whereas damage to the dorsomedial striatum renders otherwise goal-directed actions habitual. It appears, therefore, that these two regions of the striatum, or, perhaps more accurately, the distinct cortico-striatal circuits involving these regions, may compete for control of instrumental performance.

There are however, layers of complexity in attempting to understand the interaction of these apparently distinct action controllers. At one level it is clear that, at least in habitual actions, the R–O goal-directed process and the S–R habit process compete for control of performance; habitual control can apparently be immediately released and an underlying goal-directed control revealed by muscimol infusion into the DLS [93]. In fact even under normal circumstances evidence suggests that the goal-directed process can quickly suppress habitual control. This can be noticed in the everyday, e.g. while driving on a freeway when, after a period of carefree, apparently cognitively disconnected driving we see a police car approaching in the rearview mirror. Do we carry on driving in so carefree a manner? Not likely; even if we are within the speed limit and generally obeying the rules of the road, our vigilance is increased and our driving becomes more deliberated; the habit has been suppressed. Likewise, rats that are behaving habitually in extinction and so responding at a high rate on a lever trained with a sucrose solution that has subsequently been devalued, will stop responding as rapidly as non-habitual lever pressers when the lever response is punished by the actual delivery of the devalued outcome (e.g. [92,104]). The rapidity of this adjustment is, however, severely curtailed by damage or inactivation of dorsomedial striatum [77], a finding that is consistent with the argument that it is the return of control by the goal-directed system that is the source of rapid suppression of the habit in the punishment situation.

Some forms of psychopathology, most notably drug addiction, might well find their source in a defective ability to suppress habitual actions by re-engaging the goal-directed system. During the development of addiction, the pursuit of drugs of abuse rapidly becomes habitual coming under the control of internal and external states and stimuli rather than the consequences of acting [86,105,106]. It is important, however, to distinguish habitual drug seeking from other forms of habitual behavior. Under normal conditions, habit learning can be highly adaptive; habits allow us and other animals to relegate the control of routine behavioral responses to a system that uses few cognitive resources, freeing up a limited executive capacity for tasks that need greater monitoring. In contrast, habitual drug seeking is pathological; drug exposure increases the rate of acquisition of habitual actions and the influence of drug associated contexts and cues on their performance. Furthermore, despite the heavy emphasis on habit in current research on drug addiction, a distinguishing feature of habitual drug seeking is the addicts’ loss of executive control over the habit. As is commonly noted (e.g. [107]), a distinguishing feature of drug seeking is its persistence in the face of severe negative consequences (cf. DSM IV criteria for drug abuse). The compulsive pursuit of drugs can be viewed, therefore, as the product of a drug-induced increment in habit acquisition and a drug-induced decrement in the addict’s ability to exert control over the habit in the face of persistent, negative feedback. Indeed, consistent with this argument, Robinson and co-workers have reported structural changes involving a loss of dendritic spines induced by sensitization to methamphetamine in dorsomedial striatum, a key structure implicated in goal-directed action in rodents, and
a concomitant increase in spine density in dorsolateral striatum [108].

On the other hand, there is evidence that S–R and R–O learning processes cooperate in the integration of stimulus-mediated action selection with action evaluation processes during the initiation of goal-directed actions [12,58]. Perhaps the strongest evidence for cooperation of this kind comes from studies of outcome selective reinstatement in which we have found that, when previously trained on two actions for distinct outcomes, the delivery of one or other outcome after a period of extinction on the two actions results in the reinstatement of the action that, in training followed the delivery of that outcome rather than the reinstatement of the action that delivered that outcome [58]. As we have described in more detail previously [12], the ability of outcomes to exert this effect on response selection is not affected by devaluation of the reinstating outcome. Nevertheless, using a similar training situation in which outcomes were used as explicit discriminative cues for action selection, we found that these kind of stimuli can, in fact, engage an evaluative process based on the R–O association, rendering the rate or vigor of performance of an action sensitive to the current value of the outcome that the action earned during training, i.e. the rate of performance, but not the action selection, is attenuated if the outcome earned by the reinstated action is devalued ([143]; see [58] for related findings). Hence, it appears that, in the ordinary course of events, (1) a form of stimulus–response process lies at the heart of action selection, that (2) action selection causes the retrieval of the outcome associated with that action, based on the action–outcome association, and, consequently, (3) the retrieval of the incentive value of that outcome and that these later action–outcome and outcome value processes are a necessary step towards the actual performance of the action. Hence, this selection–evaluation–initiation sequence appears to require the cooperative integration of the S–R and R–O learning processes. From this perspective, therefore, choice and decision-making is an integrative process.

It is, of course, possible that both competition and cooperation between goal-directed and habit learning processes occurs but at different times or under different conditions. For example, it is possible that R–O and S–R processes ordinarily cooperate but that overtraining provides the conditions under which stimuli not only exert control over action selection but also dominate action initiation; the strength of the stimulus–response association may allow the action to be performed before it is properly evaluated, something that accords well with the notion that habitual actions are relatively impulsive. Likewise, whilst otherwise cooperative, the inhibition of habits may be a function of the quite specific conditions induced, for example, by the delivery of unexpected negative feedback. Nevertheless, the fact that both the goal-directed and habit learning systems appear to be able to function without the other suggests that their cooperation is not necessary for instrumental performance, although it may be necessary for actions to adjust normally to changes in contingency and so for instrumental performance to remain adaptive when conditions are particularly volatile.

1 It is worth noting here – as an aside – that this account overcomes some of the general problems identified with the purely habit based account of drug addiction. For example, one argument against the claim that drug addiction reflects an abnormal increment in habit learning has been based on, albeit largely anecdotal, evidence of the highly devious and nefarious strategies that addicts devise in procuring drugs. The alternative perspective proposed here sidesteps this kind of issue by emphasizing the pathological nature of the habitual control induced, not simply by an increment in habit learning but by drug-induced abnormalities in the goal-directed system with the consequent changes in goal-directed decision-making processes and in behavioral control.

3.2. Integration and interaction in the cortico-basal ganglia network

It is not clear exactly how competition and cooperation is realized in the neural networks mediating action–outcome and stimulus-response learning. Certainly, the anatomy of the cortico-basal ganglia network provides virtually limitless possibilities for convergence, divergence, integration and interaction between the complex functions that appear to be instantiated in this circuitry and, indeed, there have been many different hypotheses advanced based on this anatomy. For example, segregation of function falls naturally from the description of parallel feedback loops connecting discrete regions of the cortex with striatum, midbrain, thalamus and feeding back to their cortical origin [109,110]. Indeed, this kind of account is well suited to the suggestion that cognitive and executive functions, including goal-directed action, are mediated by the active maintenance of patterns of neural activity in different regions of prefrontal cortex [45,111–117]. Nevertheless, evidence that the prefrontal cortex plays only a time-limited role in goal-directed learning is not predicted on this account; the loop appears to be curtailed after the new strategy has been encoded.

It is, however, possible to propose alternative hypotheses to account for these data; for example, it is possible that the action–outcome association is encoded in the medial prefrontal cortex and then consolidated in the dorsomedial striatum through direct connections between the PL and the DMS that constitute the medial loop [98].2 Alternatively, it is worth noting that the description of this loop–like architecture supersedes the earlier quite attractive idea of functional integration in the striatum through the convergence of diffuse cortical regions onto a discrete striatal targets [118]. And, indeed, despite the shift in emphasis onto cortical encoding processes, there remains strong evidence for the convergence between neurons in disparate cortical areas onto single medium spiny neurons in the dorsal striatum [119]. Unfortunately, the failure to find evidence that different regions of cortex mediate a common function stands against this account. With regard to rodent instrumental conditioning, for example, evidence suggests that whereas, as described above, prelimbic prefrontal cortex is involved in encoding action–outcome associations, the orbitofrontal cortex is not but rather plays a role in the cognitive control of action selection on the basis of reward-related cues [15,120], the anterior cingulate cortex plays a role in the resolution of conflicts in action selection [143], medial agranular cortex in encoding action sequences [121] and the extensive motor and sensorimotor regions of frontal cortex appear to be primarily involved in S–R learning and, hence, in stimulus-mediated action selection. Although these findings are not consistent with the convergence theory of the cortico-striatal network, the apparent independent functions subserved by regions of prefrontal cortex in instrumental conditioning is consistent with a parallel organization of cortico-striatal circuits. The twin notions of divergence and convergence could, however, be taken to suggest that there are some functions maintained in parallel networks whereas others are mediated by converging inputs to striatum or, alternatively, that some basic independent functions are encoded in distinct striato-nigro-thalamic networks and are integrated through convergence to allow anatomically distinct parallel circuits to generate larger functional units [122].

Over and above cortico–striatal convergence, there has been considerable recent interest in the role of the midbrain dopamine

2 Although this hypothesis appears contrary to the findings of [66], note that it is focused on the acquisition of new goal-directed actions and not on the reorganization of cortex induced by 9 months of training on discrimination reversals.
projection to the striatum in the control of distinct forms of plasticity and in the transition between different forms of behavioral control. Haber et al. [123] description of a spiraling feedback network involving ventral striatum, ventral tegmental, dorsal striatum and substantia nigra has, for example, been argued to provide the basis for transitions between goal-directed and habitual processes, and some recent evidence suggests that there may be some functional relationship in this network, at least between the ventral striatum and the dorsolateral striatum [123,136]. But, of course, there are many other possible routes through which these structures might interact including projections into dorsal striatum from the ventral striato-pallido-thalamic pathway [124], through the thalamo-striatal pathway generally [125], not to mention the opportunity for integration and interaction in output regions such as the globus pallidus where collaterals from both the dorsomedial and lateral regions have been found to converge [126,127]. These aspects of the broader basal ganglia-thalamo-cortical network have not been systematically assessed functionally and constitute complex but important targets for future studies.

4. Summary and conclusion

Whatever the neural bases of the interaction between goal-directed and habitual processes turns out to be, 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 the performance of instrumental actions. 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 plays two distinct functions serving both as a reward or goal, 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 at present, and space considerations preclude a full consideration of this issue here (cf. [11,16] for reviews), but at present it appears likely that the function of parsing the outcome into both a reward and a reinforcement signal depends on the amygdala.

In recent years considerable evidence has accumulated suggesting that the basolateral amygdala plays 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 [59,73,74]. 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 [128,129], a projection that appears to be at least partly controlled by the central nucleus of the amygdala [130]. Although direct evidence that the CeN plays a role in the reinforcement signal has not yet been reported, it is known to be involved in generating general affective responses to rewarding events [144], signals associated with rewarding events [131,132] and in the control of simple stimulus–response associations, such as those involving the performance of orienting responses to stimuli associated with food [133,134,137]. By activating both the central and basolateral amygdala, therefore, a single outcome-related event could potentially exert distinct functional effects in instrumental conditioning by controlling the production of independent reward and reinforcement signals that concatenate to distinct regions of striatum to control distinct cortico-striatal circuits.

References

[26] Chu C-C, Tranel D, Damasio AR, Van Hoesen GW. The autonomic-related cor-

Please cite this article in press as: Balleine BW, et al. The integrative function of the basal ganglia in instrumental conditioning. Behav Brain Res (2008), doi:10.1016/j.bbr.2008.10.034

ARTICLE IN PRESS

Please cite this article in press as: Balleine BW, et al. The integrative function of the basal ganglia in instrumental conditioning. Behav Brain Res (2008), doi:10.1016/j.bbr.2008.10.034
Haber SN, Fudge JL, McFarland NR. Striatonigrostriatal pathways in primates.

Nakano K, Kayahara T, Tsutsumi T, Ushiro H. Neural circuits and functional

motivational control of instrumental action. Anim Learn Behav


Clasner SV, Overmier JB, Balleine BW. The role of Pavlovian cues in alcohol


Neurosci 2006;26:3805–12.

Deroche-Gamonet V, Belin D, Piazza PF. Evidence for addiction-like behavior


Jedynak JP, Uslaner JM, Esteban JA, Robinson TE. Methamphetamine-induced


Alexander GE, Delong MF, Strick PL. Parallel organisation of functionally

segregated circuits linking basal ganglia and cortex. Annu Rev Neurosci


Kelly RM, Strick PL. Macro-architecture of basal ganglia loops with the cerebral

cortex: use of rabies virus to reveal multisynaptic circuits. Prog Brain Res

2004;143.

Morgane PJ, Galler JR, Molker DJ. A review of systems and networks of the


Calabresi P, Picconi B, Tozzi A, Di Filippo M. Dopamine-mediated regulation


El-Amamy H, Holland PC. Substantia nigra pars compacta is critical to both

the acquisition and expression of learned orienting of rats. Eur J Neurosci

2006;24:270–6.

El-Amamy H, Holland PC. Dissociative effects of disconnecting amygdala cen-

tral nucleus from the ventral tegmental area or substantia nigra on learned


Calabresi P, Picconi B, Tozzi A, Di Filippo M. Dopamine-mediated regulation


Haber SN. The primate basal ganglia: parallel and integrative networks. J Chem


Han S, Holland PC, Gallagher M. Disconnection of the amygdala central

nucleus and substantia innominata/nucleus basalis disrupts increments in

conditioned stimulus processing in rats. Behav Neurosci 1999;113:

143–51.

Pisa M. Motor functions of the striatum in the rat: critical role of the lateral


Balleine BW, Dickinson A. Effect of lesions of the insular cortex on instru-

mental conditioning: Evidence for a role in incentive memory. J Neurosci


Balleine BW, Dickinson A. Effect of lesions of the insular cortex on instru-

mental conditioning: Evidence for a role in incentive memory. J Neurosci


Balleine BW, Dickinson A. Effect of lesions of the insular cortex on instru-

mental conditioning: Evidence for a role in incentive memory. J Neurosci