Chapter 8.
General discussion

Introduction

The experiments described in this thesis addressed the role played by regions of the prefrontal cortex and ventral striatum in the control of rats' behaviour by Pavlovian conditioned stimuli, and in their capacity to choose delayed reinforcement. In this concluding chapter, the findings from these experiments will first be summarized briefly. The results have already been discussed in Chapters 3–7; in this chapter, their implications will be considered in a wider context and future research directions will be suggested. The role of the ACC within its corticocortical and corticostriatal circuits will be discussed first in the light of the present data. Different theoretical views of the process of choosing between delayed rewards will then be considered, together with the neural basis of this process. Implications for theories of nucleus accumbens function will be discussed, and lastly an overview of reinforcement processes will be presented.

Summary of results

Role of the anterior cingulate cortex in Pavlovian conditioning

The ACC has previously been strongly implicated in stimulus–reinforcer learning in the rodent, in both appetitive (Bussey *et al.*, 1996; 1997a; 1997b; Parkinson *et al.*, 2000c) and aversive settings (Gabriel *et al.*, 1980a; Gabriel *et al.*, 1980b; Buchanan & Powell, 1982a; Gabriel & Orona, 1982; Gabriel *et al.*, 1991a; Gabriel *et al.*, 1991b; Gabriel, 1993; Powell *et al.*, 1994). In Chapter 3, rats with excitotoxic ACC lesions were tested on a variety of tasks to which stimulus–reinforcer learning was expected to contribute. Lesioned rats were impaired at the acquisition of autoshaping, replicating previous findings (Bussey *et al.*, 1997a; Parkinson *et al.*, 2000c), and were also impaired when the lesion was made following training. Unexpectedly, however, they were unimpaired on a number of other tasks based on Pavlovian conditioning procedures and encompassing a range of behavioural responses. ACC-lesioned rats performed normally on a simple temporally discriminated approach task, and responded normally for a conditioned reinforcer (with normal potentiation of this responding by intra-accumbens amphetamine). They also exhibited normal conditioned freezing to an aversive CS, and normal PIT. However, ACC-lesioned rats were impaired on a two-stimulus discriminated approach task (designed to capture features both of autoshaping and the conditioned approach task on which they were unimpaired), providing direct support for the hypothesis that the ACC is critical for discriminating multiple stimuli on the basis of their association with reward.

Role of the nucleus accumbens core and shell in response-specific Pavlovian–instrumental transfer

It has previously been shown that the AcbC contributes to simple PIT (Hall *et al.*, 1999). In Chapter 4, the contribution of the AcbC and AcbSh to response-specific PIT was assessed; this more complex task involves the direction of instrumental choice behaviour by noncontingently-presented Pavlovian CSs. Lesions of the AcbC impaired the response specificity of PIT (that is, the ability of the CS to influence choice behaviour) while lesions of the AcbSh impaired PIT itself. These results present problems of interpretation in the light of other studies, discussed in Chapter 4, but closely resemble the effects of AcbC

and AcbSh lesions on the effects of intra-Acb psychostimulants on responding for conditioned reinforcement (Parkinson *et al.*, 1999b), with the shell providing 'vigour' and the core 'direction' for PIT.

Behavioural tasks used to assess preference for delayed reinforcement

In Chapters 5 & 6, two tasks testing subjects' ability to choose a large, delayed reward in preference to a small but immediate reward were investigated in detail. In Chapter 5, rats were tested on a version of the adjusting-delay schedule (1987; Mazur, 1988; 1992; Wogar *et al.*, 1993b). Surprisingly, no direct evidence was found that the subjects were sensitive to the contingencies operating in this schedule, despite the use of a novel cross-correlational analysis that successfully detected such sensitivity in a range of computer-simulated subjects. For this and other reasons, this task was not pursued further. Instead, in Chapter 6, a modified version of the 'systematic' technique of Evenden & Ryan (1996) was considered. Using this task, it was demonstrated that rats were directly sensitive to the delay to reward, preferring a large reward less when it was delayed. In a detailed behavioural analysis of the task, the effects of extinction, delay omission, reversal of the pattern of delays presented to the subjects, and satiation were examined, together with the effects of cues present during the delay to reward, thereby partially characterizing the basis of normal subjects' performance. In particular, it was found that if subjects were trained with a signal or cue present during the delay to the large reward, the cue speeded learning and supported choice of the large reinforcer.

Effects of* d*-amphetamine, α-flupenthixol, and chlordiazepoxide on preference for signalled and unsignalled delayed reinforcement

In Chapter 6, groups of rats were trained on the delay-of-reinforcement choice task, with or without an explicit signal present during the delay. *d*-Amphetamine, α-flupenthixol, and chlordiazepoxide were then administered before their performance was again tested. Amphetamine enhanced preference for the large, delayed reward in the presence of the cue, at certain doses, but uniformly depressed this preference in subjects trained without the cue. This was suggested to reflect the known effect of amphetamine to enhance the efficacy of conditioned reinforcement (Hill, 1970; Robbins, 1976; Robbins, 1978; Robbins *et al.*, 1983), and may explain discrepancies in the literature regarding the effects of amphetamine on impulsive choice (Evenden & Ryan, 1996; Richards *et al.*, 1997a; 1999; Wade *et al.*, 2000). Flupenthixol, known to depress responding for conditioned reinforcement (Robbins *et al.*, 1983; Killcross *et al.*, 1997a), had cue-dependent effects consistent with this hypothesis, though it generally decreased preference for the delayed reward. The effects of chlordiazepoxide, a benzodiazepine expected not to affect conditioned reinforcement (Killcross *et al.*, 1997a), did not depend on the cue condition: chlordiazepoxide generally reduced preference for the delayed reward.

Neural basis of preference for delayed reinforcement

In Chapter 7, the same delayed-reinforcement choice task was used to assess the contribution of subregions of the ventral striatum and prefrontal cortex to preference for delayed reward. Subjects were trained on the task in the absence of explicit cues, matched to groups, and received sham surgery or lesions of the ACC, mPFC, or AcbC before being retested. ACC lesions had no effect on choice behaviour, though lesioned subjects were slower to collect the large, delayed reward. Lesions of the mPFC altered choice, but not in a manner interpretable as an altered effect of the delays. Rather, mPFC-lesioned rats exhibited a 'flattening' of the within-session shift from the large to the small reward as the large reward was progressively delayed; this was suggested to reflect a loss of session-wide temporal stimulus control. In contrast, lesions of the AcbC dramatically and persistently impaired subjects' ability to choose the delayed reward,

even though the subjects discriminated the two reinforcers. In a new, abbreviated version of the task, infusion of amphetamine into the Acb reduced subjects' preference for the delayed reward, but surprisingly did not do so in a clear dose-, delay-, or cue-dependent manner.

The results of the lesion studies reported in this thesis may be integrated into other work within this field as shown in Table 21 (overleaf).

Anterior cingulate cortex function

The relationship of the present findings to other theories of rodent and primate ACC function were discussed in Chapter 3, in which it was suggested that the rat ACC 'disambiguates' similar stimuli for its corticostriatal circuit on the basis of their differential association with reinforcement. It has been shown that the ACC–AcbC projection is necessary for rats to acquire the autoshaping task used in the present experiments (Parkinson *et al.*, 2000c). As lesions of the AcbC also impair conditioned approach in a temporally discriminated approach task (Parkinson *et al.*, 1999b), suggesting a general role for AcbC in conditioned approach, it would be predicted that ACC–AcbC disconnection would impair the acquisition of the two-stimulus temporally discriminated approach task developed in Chapter 3. This hypothesis awaits experimental test.

The ACC provides specific information to the Acb via glutamatergic projections, through which it influences response selection in conditioned approach tasks (Parkinson *et al.*, 2000c), just as the BLA appears to do for conditioned reinforcement (Burns *et al.*, 1993) and probably for PIT (Blundell & Killcross, 2000a). In all these tasks, the glutamatergic information is in some manner 'gated' or amplified by the dopaminergic innervation of the Acb, probably under the control of the CeA (Cador *et al.*, 1991; Robledo *et al.*, 1996; Hall *et al.*, 1999; Parkinson *et al.*, 2000b; Parkinson *et al.*, submitted). On the basis of other studies reviewed in Chapters 1 & 3, it is suggested that the contributions of the BLA and ACC differ in the following way: the BLA uses a CS to retrieve the motivational value of its specific US, while the ACC directs responding on the basis of the specific CS, preventing generalization to similar CSs. These suggested roles are different — the contributions of the two structures have been dissociated using autoshaping (Bussey *et al.*, 1997a; Parkinson *et al.*, 1999a) and conditioned reinforcement tasks (Chapter 3; Burns *et al.*, 1993) — but are not dissimilar, and it may be a promising area for future research to determine how these two interconnected structures communicate, and the function of that communication.

Additionally, the results of Chapter 7 provide evidence that the ACC is not simply required when behavioural tasks become 'difficult'. In the delayed reinforcement choice task, the within-session increase in the delay to the large reward causes a progressive decline in normal subjects' success at obtaining food. This can plausibly be interpreted as an increase in task difficulty, yet ACC lesions did not impair performance. As a general role for the ACC in 'task difficulty' is an untenable interpretation, further support is inferred for the specific hypothesis that *the ACC is a reinforcement learning structure involved in stimulus discrimination*. The results of Chapter 3 are also not parsimoniously explained by a deficit in *response* discrimination, as the two-stimulus temporally discriminated approach task measured exactly the same response following presentation of a CS+ or a CS–; thus, no response discrimination was required, and yet a deficit was still observed in ACC-lesioned animals.

Table 21. Summary of lesion studies concerning the major tasks used in this thesis, and related work. Results from this thesis are emboldened; a dash (–) indicates no data are available. References (* indicates studies that have not been peer-reviewed fully): (1*) Cardinal, this thesis; (2) Bussey *et al.* (1997a); (3) Parkinson *et al.* (2000c); (4) Burns *et al.* (1993); (5) Parkinson *et al.* (2000b); (6*) Everitt *et al.* (2000b); (7*) Parkinson *et al.*, submitted; (8) Cador *et al.* (1989); (9) Parkinson *et al.* (1999b); (10*) Hall *et al.* (1999); (11) Killcross *et al.* (1997b); (12) Robledo *et al.* (1996); (13*) Killcross *et al.* (1998); (14*) Coutureau *et al.* (2000); (15*) Blundell *et al.* (2000a); (16*) Dix *et al.* (2000); (17) Taylor & Robbins (1986); (18*) Corbit & Balleine (2000a); (19) Morgan & LeDoux (1995); (20) reviewed by e.g. LeDoux (2000); (21) but see Parkinson *et al.* (1999c).

Effects of excitotoxic lesions to:	**ACC**	**mPFC**	**BLA**	**CeA**	**AcbC**	**AcbSh**	**Acb DA depletion**
Approach tasks							
autoshaping (acquisition)	**impaired**[1,2,3]	normal[2]	normal[5]	impaired[5]	impaired[3]	normal[3]	impaired (severely)[7]
autoshaping (performance)	**impaired**[1]	–	–	normal[6]	impaired[1]	–	impaired (mildly)[7]
temporally discriminated approach (Pavlovian)	**normal**[1]	–	normal[8]	normal[12]	impaired[9]	normal[9]	–
discriminated approach (instrumental contingency)	–	normal[4]	impaired[4]	–	–	–	–
discriminated approach (Pavlovian, two stimuli)	**impaired**[1]	–	–	–	–	–	–
General potentiation/suppression of instrumental behaviour by a conditioned cue							
simple Pavlovian–instrumental transfer (conditioned elevation)	**normal**[1]	–	normal[10]	impaired[10,13]	impaired[10]	normal[10]	–
conditioned suppression	–	normal[14]	normal[11]	impaired[11]	normal (whole Acb lesion)[16]		–
intra-Acb amphetamine potentiation of CRf	**normal**[1]	normal[4]	present / altered[4]	impaired[12]	loss of specificity[9]	impaired[9]	impaired[17]
Directed modulation of instrumental behaviour by a conditioned cue							
conditioned punishment	–	impaired[14]	impaired[11]	normal[11]	impaired (whole Acb lesion)[16]		–
conditioned reinforcement	**normal**[1]	normal[4]	impaired[4,8,13]	normal[8,13]	normal[9]	normal[9]	normal[17]
response-specific Pavlovian–instrumental transfer	–	–	impaired (loss of specificity)[15]	–	**impaired (loss of specificity)**[1]**;** normal[18]	**impaired (loss of transfer)**[1,18]	–
Other Pavlovian conditioning procedures							
conditioned freezing to a discrete CS	**normal**[1]	normal / enhanced[19]	impaired[20]	impaired[20]	–[21]	–[21]	–[21]
Delayed reinforcement							
Ability to choose a large, delayed reinforcer over a small, immediate reinforcer	**normal**[1]	**intact, though loss of usual pattern of responding**[1]	–	–	**impaired**[1]	–	–

There have been several suggestions that ACC dysfunction is related to impulsive behaviour or over-responding (Muir *et al.*, 1996; Bussey *et al.*, 1997a; Parkinson *et al.*, 2000c). However, Chapter 7 demonstrated that ACC lesions do not induce impulsive choice, in addition to providing evidence for a behavioural dissociation between autoshaping and impulsive choice through lesion studies of the ACC and AcbC. Over-responding to a CS– in a task such as autoshaping may reflect a failure of discrimination, rather than impulsive responding. Thus, to investigate whether the ACC is truly involved in impulsivity in any way, explicit tests of motor impulsivity (such as a paced fixed consecutive number schedule, in which subjects must avoid terminating chains of responses prematurely, or a 'stop' task, in which subjects must inhibit ongoing behaviour) or reflection impulsivity (failure to acquire sufficient information to perform a task accurately) (see Evenden, 1999b) should be administered to subjects with ACC lesions.

Finally, one of the most interesting questions about the function of the ACC concerns its apparently time-limited role in behaviour (see Chapter 3, p. 113). This makes analysis of its function more difficult, as it is not presently possible to predict accurately when in the course of behavioural training the contribution of the ACC is no longer significant. As Chapter 3 also made clear, this issue touches on the present boundaries of understanding of the way in which the representations formed during Pavlovian conditioning change over time. There are several critical issues. (1) Can overtrained Pavlovian responding be considered habitual? (2) With what structures does the ACC interact during learning, and how? Candidates include the Acb, amygdala, OFC, and PCC. The ACC may do more than simply provide a flexible behavioural controller that is effective while other structures are learning more permanent representations, but it may also actively 'teach' other structures such as the PCC (Gabriel *et al.*, 1980a, p. 162; Gabriel, 1993; Freeman *et al.*, 1996; Hart *et al.*, 1997). This hypothesis provides a testable prediction concerning appetitive autoshaping: that well-learned performance will be sensitive to PCC lesions (see Chapter 3, pp. 99 / 113) even though early acquisition is not (Bussey *et al.*, 1997a).

Theories of learning and choice with delayed reward

Two broad approaches to choice behaviour will be summarized, and a synthesis offered.

Model 1 (informed choice). According to this model, subjects make prospective choices between alternatives based on full knowledge of the response–outcome contingencies and of the value of each outcome. These choices represent goal-directed actions. Subjects' sensitivity to delay in choice tasks is therefore a consequence of time discounting of the perceived (prospective) value of the delayed reward.

This model is necessarily applicable only to fully-trained subjects — subjects who have learned the instrumental contingencies. It may be particularly applicable when humans are offered explicit hypothetical choices ('would you prefer $800 now, or $1000 in a year?'; Rachlin *et al.*, 1991; Myerson & Green, 1995).

As the contingencies cannot be offered 'pre-packaged' to experimental animals through language, such subjects must be trained through direct experience of the rewards in the experimental situation. This introduces the complication that delays to reinforcement can affect operant and discrimination learning (reviewed in Chapter 1), so care is typically taken by experimenters to ensure subjects are 'well trained'. Slow acquisition of delay sensitivity must be attributed to difficulties in learning the instrumental contingencies across a delay and / or learning the appropriate incentive value of delayed reward through experience of waiting. In tasks where the delay is systematically and predictably varied, as in Chapters 6 & 7, learning may also be slowed by the requirement to learn S^Ds predicting the delay contingency currently in

force. Thus, this model is inherently an incomplete description of the effects of delayed reinforcement, as it does not deal with the effects of delays on learning.

Model 2 (associative response strength). According to an extreme form of this model, based on simple S–R theory (Thorndike, 1911; Grindley, 1932; Guthrie, 1935; Hull, 1943), rats' choice behaviour reflects differential reinforcement of stimulus–response habits. The change in associative strength is some function of reward magnitude multiplied by the time-discounted 'trace strength' of the preceding response. Choice is determined by some process of competition between the available responses (e.g. the principles of matching; Herrnstein, 1970; de Villiers & Herrnstein, 1976). Choice is therefore 'retrospective' in a sense, as preference for a particular alternative depends upon prior experience of that alternative, and time discounting reflects the decay of the traces available to be associated with reward. A similar model, after Grice (1948), may be constructed in which animals respond for immediate conditioned reinforcement (by goal-directed behaviour or S–R habit) and the acquisition of associations between a chain of stimuli and eventual reward accounts for the observed phenomenon of temporal discounting, by similar mechanisms.

The S–R view accounts for some of the theoretical appeal of exponential temporal discounting models. In exponential decay, at any one moment in time the trace strength of a response follows directly from the trace strength at the previous instant (if x_t is the trace strength at time t and A is the starting value, then $x_t = \mathrm{A}e^{-kt}$ and $x_{t+1} = e^{-k}x_t$). In contrast, in the hyperbolic discounting model and all others in which preference reversal occurs, the strength of the trace at any one moment cannot be calculated in such a manner: information about the absolute time since the response must be available. (This may be clearly illustrated by the preference reversal graph shown in Chapter 1, p. 58; if two such decay curves cross, then an observer travelling along one curve cannot know at the crossover point whether its own curve is the recent, rapidly-decaying trace, or the older, slowly-decaying trace, without further information — namely the time since the response or its starting strength.) This process does not model 'mnemonic delay' in any clear way. Thus, the empirical observation of hyperbolic discounting specifies the information that must be available to the subject at any one moment in time; in the context of 'retrospective' choice, this constrains the possible underlying psychological mechanisms, and there is no obvious candidate within the S–R model.

While S–R models can account for effects of delays on learning as well as choice, they do not take into account the fact that goal-directed actions contribute to choice in rats (Dickinson, 1994) and would clearly not provide a satisfactory account of human choice (cf. Ainslie, 1975; Rachlin *et al.*, 1991; Myerson & Green, 1995).

Model 3 (composite). A multifactorial model is therefore suggested, based on that of Dickinson (1994). The 'response strength' of any behaviour is governed by (1) goal-directed action (Dickinson & Balleine, 1994), in which knowledge of the instrumental contingency combines with the incentive value of the expected outcome; (2) stimulus–response habits, which gain strength slowly with the number of reinforcers presented (Dickinson *et al.*, 1995); and (3) PIT, mediated by the Pavlovian association between contextual, discriminative, or other conditioned stimuli and the outcome of the instrumental action. Ordinarily, behaviour conforming to the matching law and to hyperbolic temporal discounting is seen as a product of these processes. Delayed reinforcement may act (a) to impair learning of the instrumental contingency (Dickinson *et al.*, 1992); (b) to reduce the incentive value of the delayed reward, as speculated by many models; (c) to reduce the reinforcement of stimulus–response habits; and (d) to reduce the Pavlovian association between stimuli present at the moment of action and the ultimate reinforcer.

This model makes several predictions. Firstly, manipulations of components of this composite behaviour should affect choice. For example, manipulations of the association between cues immediately con-

sequent on choice and the outcome (e.g. presence of absence of a cue bridging the delay) should affect choice independently of the actual delay to reinforcement, a prediction not made by Kacelnik's (1997) normative model of hyperbolic discounting, but one supported by the results of Chapter 6. Secondly, pharmacological and neural manipulations known to dissociate these processes should also be capable of affecting choice.

This view is obviously compatible with mathematical models of temporal discounting, but interprets the discount function as the sum of the contributions of several processes operating in any one situation. Similar composite models have been offered before (a casual example is Pinker, 1997, pp. 395–396), though with a different decomposition of the processes contributing to choice (e.g. distinct contributions of conditioned and primary reinforcement to response strength; Killeen & Fetterman, 1988, pp. 287–289). One interesting challenge may be to establish what processes contribute most significantly to choice of a reinforcer at different delays. Consider an obvious hypothesis: instrumental incentive value in the rat depends upon declarative knowledge, as discussed in Chapter 1 (p. 23), and in this way is analogous to human hypothetical choices. Thus it may be that when reward is extremely delayed (as in some human experiments), only instrumental incentive value is important (as delay $d \rightarrow \infty$, total value $V \rightarrow V_{\text{instrumental}}$). When a dieting human calmly decides to abstain from chocolate cake and the dessert trolley is then pushed under his nose, it would not be expected from the rat literature that the instrumental incentive value of chocolate cake suddenly changes — after all, the subject's underlying motivational state of hunger (or lack of it) has not altered. However, alternative, possibly Pavlovian motivational processes may create an extra boost to the value of the cake (observed as a tendency to choose the cake), which is now immediately available (as $d \rightarrow 0$, $V_{\text{cake–other}}$ increases dramatically). The net value function ($V = V_{\text{cake–instrumental}} + V_{\text{cake–other}}$) could then exhibit preference reversal, leading our diner to succumb and choose the immediate reinforcer. This illustrates but one possible scenario. Nevertheless, if different processes do contribute at different delays, there would be important implications for our understanding of individual differences in impulsive choice.

Devaluation of the delayed reinforcer may yield clues concerning this suggestion in experimental animals. For example, if devaluation led to a fall in preference for a delayed reward (relative to a non-devalued, immediate alternative) when subjects were tested in extinction, this would suggest that instrumental incentive processes were prominent contributors to the overall 'value' of the delayed reward, while failure to observe this would suggest habitual responding. If instrumental incentive processes contribute more to the value of an immediate reinforcer than to that of a delayed reinforcer, and the two reinforcers were of the same foodstuff, devaluation might even lead to an increase in preference for the delayed reinforcer (assuming the subjects did not simply cease responding). Of course, a practical problem might be that the use of a discrete-trial schedule may encourage stimulus-bound responding in a way that free-operant schedules do not, while providing frequent choices may discourage habit formation.

Theories of nucleus accumbens function and the neural basis of delayed reward

1. The striatum as a switching device

A quarter of a century ago, Lyon and Robbins (1975) hypothesized a behavioural switching mechanism based on the dopaminergic innervation of the striatum. This concept has evolved (Robbins & Sahakian, 1983; Robbins *et al.*, 1990b); Redgrave *et al.* (1999a) recently reviewed and extended theories of the basal ganglia as a central behavioural switching mechanism (Lyon & Robbins, 1975; Cools, 1980; Dunnett & Iversen, 1982; Jaspers *et al.*, 1984; Redgrave *et al.*, 1999a), which provides a useful framework within

which to discuss the present data (see also Parkinson *et al.*, 2000a). According to this theory, the striatum selects responses in the cortical structures to which it is connected, by disinhibiting one 'channel' passing through it, and using a winner-take-all system to ensure that only a single channel is active. Superimposed upon this picture may be a hierarchy: whilst the motor loop of the dorsal striatum switches between incompatible commands to the musculature, the limbic loop (ventral striatum) may operate at a higher level to switch between different overall behavioural strategies. The concept of hierarchical switching is illustrated in Figure 100 (with detail of the ventral striatal circuit in Figure 101). This mechanism is an efficient way to resolve conflicts over access to limited motivational, cognitive and motor resources (Redgrave *et al.*, 1999a).

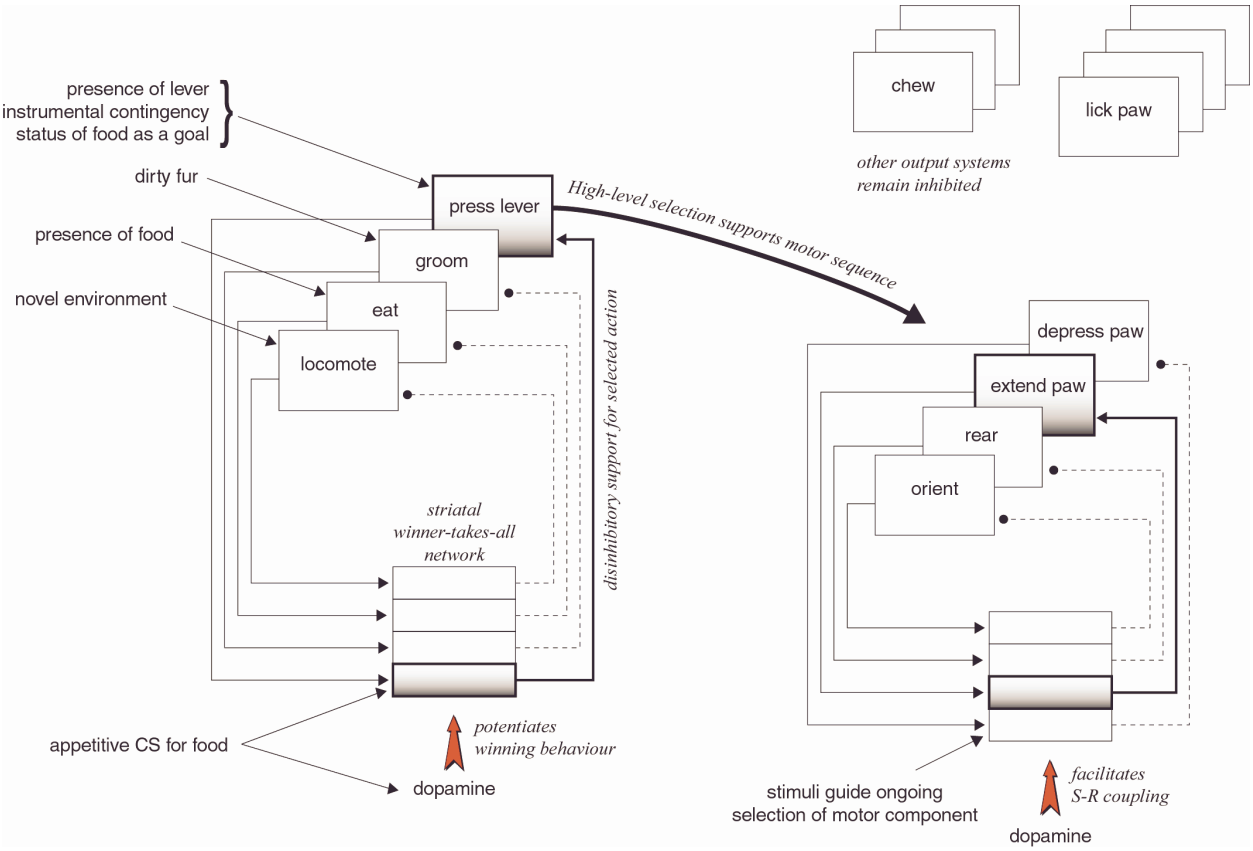

Figure 100 illustrates concepts of central switching mechanisms and hierarchies of behaviour. The left-hand circuit, representing the limbic corticostriatal loop, influences the selection of complex behaviours on the basis of conditioned motivational stimuli. The right-hand circuit, representing the motor corticostriatal loop, selects motor responses on the basis of environmental stimuli in an S–R fashion. The interaction between the circuits represents the hierarchy of behaviour: motor components can only be selected when they are part of the chosen higher-level behaviour.

Striatal circuitry is consistent with this hypothesis. Striatal medium spiny neurons are well suited by their connectivity and electrophysiology to act as pattern detectors: they are bistable, receive a highly convergent projection from the cortex and require cortical input to enter the active ('up') state. They are therefore suited to 'registering' patterns of cortical input (see Houk & Wise, 1995; Wilson, 1995) and appear to do so (Schultz *et al.*, 1995a). More controversially, they may receive a 'teaching signal' to influence future recognition of cortical patterns of activity, discussed later. A caveat is that the neostriatum is only able to discriminate cortical input patterns that are linearly separable, as it is equivalent to a single-layer network (Wickens & Kötter, 1995), and its discriminative ability is further limited by the fact that direct

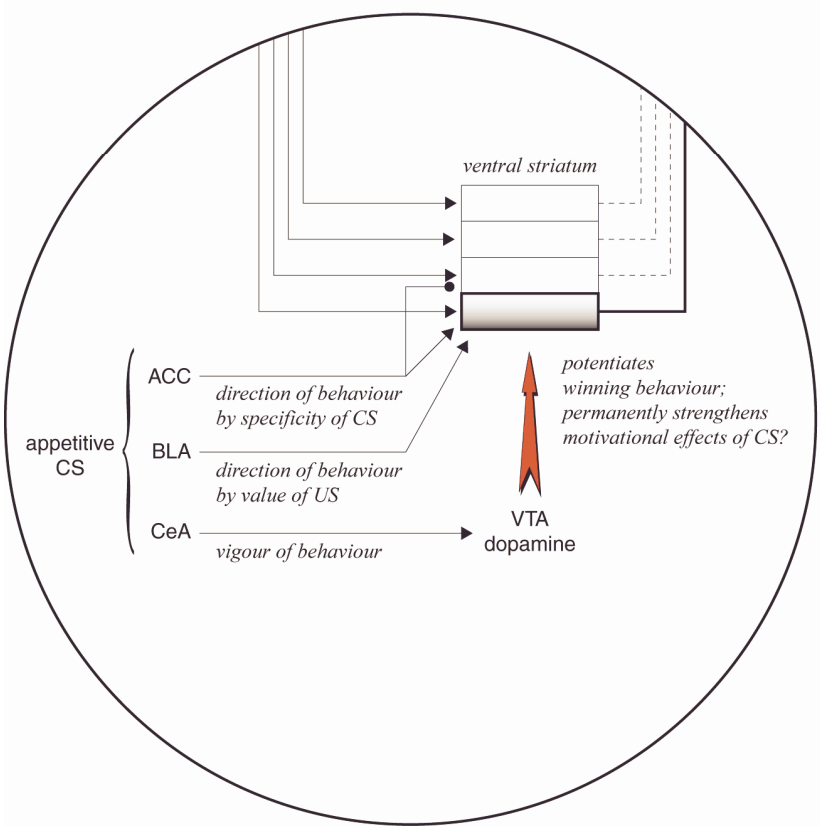

Figure 101. Speculative view of influences mediated through the ventral striatum ('close-up' of the left-hand circuit of Figure 100). Information about CSs may influence the Acb in several ways. The BLA is implicated in the retrieval of the current value of the US (see Everitt *et al.*, 2000a), and the modulation of choice behaviour; for example, it is required for CRf (Cador *et al.*, 1989; Burns *et al.*, 1993) and response-specific PIT (Blundell & Killcross, 2000a); the latter requires the AcbC (Chapter 4). Among its other roles, the CeA projects to the VTA and is required for the invigorating effect of CSs on instrumental responding (in tasks such as simple PIT; Hall *et al.*, 1999) and on locomotor approach (in autoshaping; Parkinson *et al.*, 2000b), probably via its effects on Acb DA (Hall *et al.*, 1999; Parkinson *et al.*, submitted). The ACC appears to be required to discriminate similar CSs on the basis of their association with reward (Chapter 3), preventing inappropriate responses through its projection to the Acb (Parkinson *et al.*, 2000c). Ventral striatal DA may enhance ongoing responding, but may also 'teach' the striatum; it is speculated that this can lead to a permanent enhancement in the motivational impact or salience of a CS, or its ability to induce certain patterns of motivated response (cf. Robinson & Berridge, 1993) (and see text, pp. 243 / 245).

corticostriatal projections are glutamatergic and excitatory. Within the major corticostriatal loops (skeletomotor, oculomotor, 'cognitive', and limbic), there are parallel channels: circuits that maintain a degree of functional segregation (Alexander *et al.*, 1986) and that may compete for output (Deniau *et al.*, 1982). Striatal output circuitry may operate on a disinhibitory principle: GABAergic neurons in the globus pallidus and SNr tonically inhibit thalamocortical circuits, and activity in GABAergic striatal neurons can inhibit globus pallidus / SNr neurons, disinhibiting the cortex (see Alexander & Crutcher, 1990; Chevalier & Deniau, 1990). This disinhibition does not itself trigger behaviour, but *permits* it (reviewed by Chevalier & Deniau, 1990), as the striatum does not generate simple behaviour patterns, but chooses and / or links them. This concept has been well illustrated by studies of grooming in rats; small (~1 mm^3) excitotoxic lesions of the dorsal striatum can impair the *sequence* of grooming behaviour without affecting the rat's capacity to emit any component of the sequence (see e.g. Aldridge *et al.*, 1993; Cromwell & Berridge, 1996). Those studies that have explicitly looked at switching are also consistent with this hypothesis; thus Acb lesions have been shown to impair 'strategy switching' in a reversal situation (Reading & Dunnett, 1991), though not on all tasks (Stern & Passingham, 1995). Chapter 4 provided further evidence for a role of the ventral striatum in the direction of ongoing behaviour by conditioned stimuli, distinguishing in addition between the AcbSh, which provided the 'vigour' for Pavlovian–instrumental transfer, and the AcbC, which provided the direction or response specificity. The manner in which the AcbC and AcbSh interact in PIT is not yet clear, as for CRf (see Chapter 1), but the 'vigour' / 'direction' hypothesis is consistent with theories postulating a hierarchy even within the ventral striatum, from shell to core (e.g. Haber *et al.*, 2000).

2. Acute modulation of striatal function by dopamine

Whilst glutamatergic afferents to the striatum constitute high-bandwidth pathways, capable of carrying a large amount of information, the dopaminergic input is a low-bandwidth pathway (Schultz, 1994; Mirenowicz & Schultz, 1996; Zoli *et al.*, 1998), consistent with a role in modulating other information passing through the striatum (the functions of striatal dopamine are of necessity entirely constrained by the underlying function of the striatum). Direct evidence for such a modulatory role is provided by the conditioned reinforcement paradigm (Taylor & Robbins, 1984): infusion of dopaminergic agonists into the Acb increases the rate (i.e. the momentary probability) of responding for a conditioned reinforcer, but can only 'amplify' this effect of conditioned reinforcers when information about them is arriving via glutamatergic afferents, in this case from the BLA (Cador *et al.*, 1989; Burns *et al.*, 1993).

At a cellular level in the striatum, dopamine probably focuses activity by increasing output from the most active medium spiny neurons (which are in the minority) and decreasing output from the less active cells (see Grace, 1987; Wickens & Kötter, 1995). This is mirrored at a behavioural level; increasing doses of dopamine agonists produce higher rates of activity in more and more limited categories of response (Lyon & Robbins, 1975) until stereotypy ensues.

The differences in the functions of dopamine in the dorsal and ventral striatum (reviewed by Robbins & Everitt, 1992) can then be viewed as a common action of dopamine on striatal circuits that switch different aspects of behaviour (cf. Alexander *et al.*, 1986). In the dorsal striatum, dopaminergic agonists alter the relative probability of simple motor acts, leading to stereotypy at high doses. Antagonists and dopamine depletion prevent relevant stimuli from eliciting simple motor responses, including consummatory responses; the spectrum is from a slowed response to akinesia. Similarly, dopamine depletion of the dorsal striatum impairs learning and performance of tasks based on stimulus–response decision rules (Robbins *et al.*, 1990a). As would be predicted from the corticostriatal loop account, cognitive aspects of stimulus–response coupling, such as the establishment and maintenance of an attentional or response 'set', are probably also impaired by dorsal striatal dopamine depletion (see Marsden, 1992; Robbins & Everitt, 1992, p. 122). This description emphasizes the role of the striatum as a device that selects behavioural output *in appropriate stimulus situations.*

In the ventral striatum, dopamine agonists and antagonists similarly increase or decrease the probability of stimuli affecting ongoing behaviour, but the behaviour so altered is qualitatively different. When intra-Acb amphetamine is given to rats responding for CRf, the response that is potentiated is a complex motor act, arbitrarily chosen by the experimenter, and induced by a process of conditioned motivation. The ventral striatum also mediates motivational influences on locomotion and on preparatory aspects of behaviour (Robbins & Everitt, 1992). Switching between complex behaviours is itself reduced by dopamine depletion or antagonist injection into the Acb (Koob *et al.*, 1978; Robbins & Koob, 1980; Evenden & Carli, 1985; Bakshi & Kelley, 1991).

3. The striatum and learning

The question of whether the striatum itself is involved in learning is controversial. If the switching hypothesis is correct, then striatal learning would manifest itself as a permanent change in the probability of a particular cortical pattern or behaviour being disinhibited by the striatum, given a certain pattern of inputs. Such a mechanism would also be capable of learning motor sequences. As discussed in Chapter 1, a role for the basal ganglia in habit formation was originally suggested by Mishkin *et al.* (1984), who saw a habit as a direct stimulus–response association that was learned slowly but was stable. Much of the subsequent work on this issue has proved controversial (see Wise, 1996; Wise *et al.*, 1996; White, 1997),

though Packard & McGaugh (1996) have provided good evidence for a long term change in behaviour that is dependent on the striatum. In their experiment, described in Chapter 1 (p. 46), rats were trained in a T-maze with one arm consistently baited. This task is soluble by repeating the reinforced response, or by approaching the place where food was found (a 'place response'), and these alternatives were distinguished by letting the rat approach the choice point from the opposite direction. After 8 days of training, most rats made place responses, which depended on the function of the dorsal hippocampus but not of the dorsolateral caudate nucleus. After 16 days of training, however, most rats instead made the motor response that had been reinforced. Inactivating the caudate with lidocaine eliminated this tendency and reinstated place responding, whilst inactivation of the hippocampus had no effect. Therefore, in this task, development of a stimulus-to-motor response mapping takes place slowly during reinforced training and comes to dominate behaviour, and its performance depends on the caudate nucleus. However, this response has not yet been characterized as a habit by reinforcer devaluation techniques; similarly, it is not clear from this type of experiment whether the caudate itself is the critical site of plasticity or is merely involved in behavioural expression of the response.

4. Dopaminergic effects on striatal learning; implications for addiction

Dopamine has been widely suggested to affect learning by effects exerted within the striatum. At a cellular level, dopamine can mediate heterosynaptic plasticity in the striatum (reviewed by Wickens & Kötter, 1995): pre- and postsynaptic activity in the corticostriatal pathway produces long-term depression (Calabresi *et al*., 1992) but phasic dopamine may reverse this, producing a potentiation (Wickens *et al*., 1996) (though see Pennartz *et al*., 1993). Single-cell recording has shown that dopaminergic neurons of the SNc / VTA respond to unpredicted rewards; with training, this response transfers to stimuli predictive of rewards (Schultz *et al*., 1993; Mirenowicz & Schultz, 1994; Mirenowicz & Schultz, 1996). Based on the response properties of midbrain dopamine neurons, computational neuroscientists have suggested that by signalling reward prediction errors, dopamine acts as a teaching signal for striatal learning (Houk *et al.*, 1995; Montague *et al*., 1996; Schultz *et al.*, 1997) in a system based upon temporal difference (TD) learning (Sutton, 1988), with dopamine increasing the probability of repeating responses that lead to reward. It would certainly be maladaptive to develop inflexible, habitual behaviour if such learning were not guided by a signal at least correlated with reinforcement, and the dopamine signal fulfils this property.

While the suggestion that dopamine acts as a teaching signal is controversial (e.g. Pennartz, 1995; Redgrave *et al*., 1999b) — for example, many effects of dopaminergic manipulations are interpretable as effects on attentional or response switching — there is some behavioural evidence for dopaminergic consolidation of S–R learning. The 'win-stay' radial maze task may be solved by a stimulus–response rule, as approach to an illuminated arm is always rewarded. Performance on this task is also sensitive to caudate lesions (Packard *et al*., 1989) and improved by post-training injections of dopamine agonists into the caudate (Packard & White, 1991). These effects are neurally and behaviourally specific: caudate manipulations had no effect on a 'win-shift' task in the same apparatus, and were doubly dissociated from the effects of lesions of the hippocampus or post-training hippocampal injections of dopaminergic agonists. Post-training microinjections represent a critical experimental test for the demonstration of task consolidation, as they cannot affect task performance. However, the task cannot be characterized as a stimulus–response habit as clearly as the T-maze task.

Does ventral striatal dopamine consolidate learning of a stimulus–motivation mapping in a similar manner? At present, this is an unanswered question. Unpublished observations from our laboratory indicate that rats responding for a conditioned reinforcer in extinction under saline conditions respond more if they have previously responded with intra-accumbens amphetamine, which is contrary to the general ten-

dency for responding to extinguish (R.N. Cardinal, T.W. Robbins and B.J. Everitt, unpublished observations). However, these data are confounded by response generalization effects (the possibility that the rats responded more simply because they had a history of high responding in the same environment). Post-training injections of the dopamine D_2 antagonist sulpiride into the Acb have been shown to impair water-maze performance (Setlow & McGaugh, 1998; 1999), but the theoretical basis of this task is not clear.

Since the dorsal striatum is involved in the development of stimulus–response habits (Reading *et al.*, 1991; Packard & McGaugh, 1996), whilst the ventral striatum is involved in motivational processes (Robbins & Everitt, 1992), a qualitative difference may exist between the two. However, if the two structures (at least, the dorsal striatum and the AcbC) perform similar functions at a neural level, then a direct comparison may be fruitful. An S–R habit may be defined as the production of a motor response with a fixed probability given a set of stimuli; that is, a simple and inflexible input / output mapping. Habits are also learned slowly. But if the striatum subserves S–R habits, then the stimulus is whatever cortical inputs arrive at a striatal segment, which depends upon the corticostriatal loop of which it is part, and the response is the pattern that the striatum consequently induces in the structures to which it projects. For the ventral striatum, the equivalent habit would be the *inflexible generation of a motivational effect* in a particular context.

Such 'motivational habits' may be of critical importance in the phenomenon of drug addiction (Figure 102). Compulsive drug use is characterized by behaviour that is inflexible, for it persists despite considerable cost to the addict and may become dissociated from subjective measures of drug value (Robinson & Berridge, 1993), may be elicited by specific environmental stimuli (O'Brien *et al.*, 1986), and yet involves complex, goal-directed behaviours for procuring and self-administering a drug. In a behavioural hierarchy, inappropriate reinforcement of low-level behaviours may be of trivial consequence whereas drug-induced reinforcement of a motivational process that has flexible cognitive and motor systems at its disposal may be far more destructive.

A critical question regarding the neural basis of addiction is what differentiates the effects of abused drugs from the effects of natural reinforcers, and whether this is a qualitative or a quantitative difference. In addition to the ability of drugs of abuse to activate DA systems more consistently than food reinforcers (see Di Chiara, 1998), recent evidence suggests that one effect unique to such drugs may be *sensitization*, the phenomenon by which repeated drug administration leads to an enhanced response to the drug (for reviews, see Robinson & Berridge, 1993; Altman *et al.*, 1996, pp. 302–304). Sensitization to amphetamine is induced via the drug's effects on VTA cell bodies and is expressed as hypersensitivity to amphetamine at dopaminergic terminals in the Acb (Cador *et al.*, 1995; see also Stephens, 1995; Kalivas *et al.*, 1998; Mead & Stephens, 1998). The Pavlovian motivational process suggested to be subserved by ventral striatal dopamine (as discussed in Chapter 1, p. 46) and termed *incentive salience* or 'wanting' by Robinson & Berridge (Robinson & Berridge, 1993; Berridge & Robinson, 1998) has been suggested to sensitize (Robinson & Berridge, 1993); 'incentive sensitization' may be an important contributor to addiction. The potential link to PIT and amphetamine potentiation of CRf is clear. PIT may be an important basis for conditioned reinforcement (see Chapters 1 & 4, pp. 31 / 50 / 146) and intra-accumbens amphetamine potentiates PIT (Wyvell & Berridge, 2000, using a food US). Sensitization interacts with these Acb-dependent processes: amphetamine sensitization leads to enhanced conditioned approach and conditioned increases in amygdala dopamine in response to a CS (Harmer & Phillips, 1999), while repeated cocaine administration sensitizes the response to intra-accumbens amphetamine when responding for CRf (Taylor & Horger, 1999, using a water US). It seems likely that PIT would sensitize in a similar way; this will be an important suggestion to test, and if confirmed it will be particularly interesting to test whether psy-

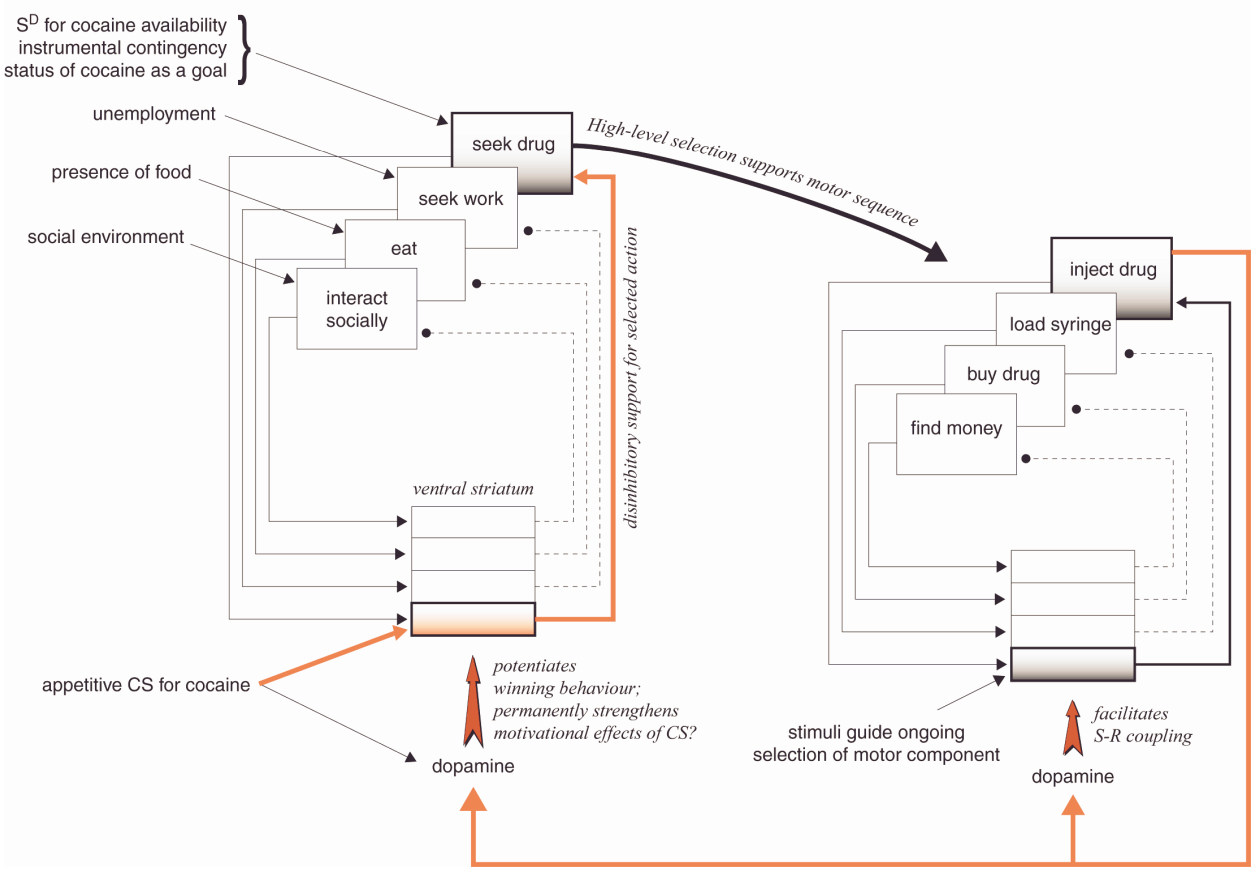

Figure 102. Highly speculative version of Figure 100 illustrating in red the particular problem posed by drugs of abuse, such as cocaine. While goal-directed behaviour may lead an individual to take such drugs (top left), just as it leads to other goals, drugs of abuse are particularly powerful at activating dopamine systems. It is possible that ventral striatal DA permanently enhances the motivational impact of Pavlovian CSs (bottom left), making those CSs potent at influencing instrumental behaviour (cf. Robinson & Berridge, 1993). If this DA system were abnormally enhanced, the Pavlovian CS might become capable of triggering complex drug-seeking behaviour even if the drug did not have high instrumental incentive value — a motivational habit.

chostimulant sensitization enhances PIT regardless of the US (for example, whether repeated noncontingent amphetamine would enhance PIT using a food US), as suggested by the comparison between Taylor & Horger (1999) and Wyvell & Berridge (2000), or whether such sensitization would predominantly affect PIT for psychostimulant USs. PIT may be especially important in addiction (with potential roles in acquisition, maintenance, and cue-induced relapse; see e.g. Tiffany & Drobes, 1990; Gawin, 1991; O'Brien *et al.*, 1998) as it represents a mechanism by which uncontrolled (noncontingent) stimuli can radically affect goal-directed responding.

5. Incorporation of the present findings relating to delayed reinforcement

The finding that the AcbC is critical for choosing delayed reward is intriguing and novel, but the psychological basis of this effect is not yet clear. As discussed in Chapter 7 (p. 226), it will be important to establish whether this deficit is entirely due to a difference in the perception of reward magnitude in AcbC-lesioned rats, or whether a specific delay-dependent deficit exists. Furthermore, as discussed above (p. 239), the psychological processes contributing to choice at different delays are not well understood at present. It is therefore unclear whether the deficit in AcbC-lesioned rats can be interpreted entirely within the framework of ventral striatal function reviewed here and by Parkinson *et al.* (2000a). If it can be shown behaviourally that Pavlovian conditioned motivation is a major contributor to preference for delayed rewards, this may be accomplished. However, as the task used in Chapter 7 had no explicit cues signalling

the delayed reward, this explanation appears specious and *post hoc*, and theories of Acb function may have to be extended (as discussed in Chapter 7, p. 230) to accommodate the novel finding.

It will therefore be important to assess the roles of the AcbSh and Acb DA in preference for delayed reward, using excitotoxic and DA-depleting lesions respectively, particularly given the evidence from studies of ADHD and animal models thereof implicating ventral striatal dopamine in the pathogenesis of impulsive choice (reviewed in Chapters 6 & 7).

To examine the possibility that the AcbC plays a wider role in learning across delays, it will be interesting to establish whether AcbC-lesioned rats are impaired at instrumental learning when the reinforcer is delayed (in the absence of an immediately-available alternative). Demonstration of such an impairment would tend to support the view that AcbC-lesioned rats prefer immediate reinforcement because they have difficulty learning that the alternative choice leads to reinforcement at all, while failure to find an impairment would suggest that AcbC-lesioned rats are 'aware of their options' even as they choose the immediate reinforcer. More generally still, it remains to be established whether AcbC lesions impair *Pavlovian* conditioning when the CS–US interval is long (trace conditioning). Additionally, trace conditioning may be less effective than conditioning with a short CS–US interval because trace conditioning promotes conditioning to other stimuli occurring during the interval, including contextual stimuli (see Dickinson, 1980, pp. 61–70; Mackintosh, 1983, pp. 202–210). AcbC lesions have been shown to impair conditioning to discrete cues but enhance contextual conditioning in a lick suppression task (Parkinson *et al.*, 1999c). The possibility may be entertained that hippocampal lesions (which disrupt contextual conditioning in the same task; Selden *et al.*, 1991) might promote responding for delayed reinforcement by reducing contextual overshadowing in instrumental learning.

Finally, even if AcbC lesions are shown to cause a purely delay-dependent (rather than reward magnitude-dependent) deficit, it would be extremely unusual for striatal lesions to impair a behaviour not impaired by lesions to its afferents (other than behavioural sequencing and switching, as discussed above, pp. 241–243). Thus, before ascribing a specific delay-dependent function to the Acb, it must be shown that lesions of the afferents to the Acb do not produce the same deficit. This work was begun in the present thesis with the demonstration that ACC and mPFC lesions do not impair rats' ability to choose a delayed reward, but the effect of lesions to other glutamatergic afferents such as the BLA, the subiculum, and the orbitofrontal cortex are unknown, as are the effects of manipulations of the DA and 5-HT innervation of the Acb. The task developed by Evenden & Ryan (1996) has proved very useful in this field of study. It has now been successfully applied to pharmacological, behavioural, and lesion studies of delayed reinforcement, and will likely prove a good starting point for future work to elucidate further the neural circuit responsible for the important ability of animals to gain reinforcement, even when it is delayed.

Reinforcement learning in the brain: an integrative view

Reinforcement is not unitary. As reviewed in Chapter 1, Pavlovian conditioning creates multiple representations. Their neural bases are gradually becoming clear. These include CS–US(sensory) or S–S associations, required for sensory preconditioning and dependent at least in part on the perirhinal cortex for visual stimuli and on the gustatory neocortex for food USs; CS–US(motivational) associations, suggested to depend on the BLA; direct CS–affect associations, which are responsible for transreinforcer blocking and are poorly understood; and CS–response associations, whose neural basis depends on the specific response (being cerebellum-dependent in the case of discrete skeletomotor CRs, and CeA-dependent in the case of several others such as conditioned suppression and PIT). Learning theorists discovered the

existence of these multiple representations and learning theory has dramatically enhanced neurobiological studies of conditioning, but sometimes neural dissociations within conditioning have been found that were not predicted by learning theory (e.g. Steinmetz, 2000); in these situations, neurobiology can inform learning theory. When considering the manner in which representations change with training, and how these representations are formed and interact across widely distributed neural systems, neither behavioural studies nor biology have provided clear answers and much work remains to be done.

Other structures contribute to instrumental conditioning (also reviewed in Chapter 1), which also creates multiple representations and can be heavily influenced by Pavlovian conditioning procedures. The prefrontal (prelimbic) cortex is critical for the perception of instrumental contingencies, while gustatory neocortex also has a role in recalling the instrumental incentive values of foodstuffs. It is not yet known how either structure acquires or represents this information, or how they interact with other representations of stimulus and reward value such as those in the amygdala and orbitofrontal cortex. It seems likely that the dorsal striatum contributes in some way to the acquisition of S–R responding, but this requires definitive proof. The nucleus accumbens was accurately described by Mogenson *et al.* (1980) as a limbic–motor interface, but may also be considered a Pavlovian–instrumental interface; it is a critical site for the motivational and directional impact of Pavlovian CSs on instrumental responding and locomotor approach. This multiplicity of representations should guide modelling studies: simple computational models of reinforcement learning, typically S–R in nature, may provide useful information regarding the principles upon which S–R systems can operate, but are often inadequate for describing simple instrumental behaviour in rats. At least some of the processes governing instrumental responding are based on declarative knowledge that is akin to symbolic processing, and yet these complex representations are known to interact with each other and with basic motivational states. Understanding this interface, and with it the nature of neural representations themselves, is one of the greatest challenges for neurobiology.

This thesis has not answered or even addressed the vast majority of these questions, but it has provided evidence that the anterior cingulate cortex makes a discriminative contribution to Pavlovian conditioning; it has elucidated further the manner in which the nucleus accumbens core and shell mediate the impact of Pavlovian CSs on instrumental responding; it has demonstrated an interaction between Pavlovian CSs and the effects of psychostimulant drugs on choice of delayed reinforcement, and it has demonstrated that the nucleus accumbens core is a critical part of the neural circuitry mediating the effects of delayed reinforcement on instrumental responding.