Chapter 1: Introduction

1.1 OVERVIEW

The problems of learning with delayed and uncertain reinforcement, and of choosing delayed or uncertain rewards, are interesting from a theoretical and a practical perspective.

The theoretical perspective concerns the neural mechanisms of learning and choice. Animals act to obtain rewards including food, shelter, and sex. Sometimes, their actions are rewarded or reinforced immediately, but often this is not the case. Often, natural reinforcers follow the action that obtains them by a delay, even if it is short; to be successful, animals must learn to bridge these delays and act on the basis of delayed reinforcement. They may also profit by choosing delayed reinforcers over more immediate reinforcers, if the delayed reinforcers are sufficiently large. Likewise, animals that can act despite uncertainty as to what the future holds and can calculate risk appropriately are placed at a competitive advantage.

The practical perspective concerns individual variation in sensitivity to delayed and uncertain reinforcement. Individuals differ in their ability to choose delayed rewards. Self-controlled individuals are strongly influenced by delayed reinforcement, and choose large, delayed rewards in preference to small, immediate rewards; in contrast, individuals who are relatively insensitive to delayed reinforcement choose impulsively, preferring the immediate, smaller reward in this situation (Ainslie, 1975). Impulsivity has long been recognized as a normal human characteristic (Aristotle, 350 BC / 1925) and in some circumstances it may be beneficial (Evenden, 1999b), but impulsive choice contributes to deleterious states such as drug addiction (Poulos *et al.*, 1995; Heyman, 1996; Bickel *et al.*, 1999; Evenden, 1999a; Mitchell, 1999) and attention-deficit/hyperactivity disorder (ADHD) (Sagvolden *et al.*, 1998; Sagvolden & Sergeant, 1998). Similar individual differences in the tendency to work for uncertain rewards may contribute to personality traits such as venturesomeness (Eysenck & Eysenck, 1977; Evenden, 1999a), but risk taking is another aspect of impulsivity (Daruna & Barnes, 1993; Eysenck, 1993; Evenden, 1999a) and is a feature of a number of psychiatric disorders, including pathological gambling and certain personality disorders (Roy *et al.*, 1989; Coccaro & Siever, 1995; APA, 2000; Holt *et al.*, 2003).

In this chapter, top-down (behavioural economic) and bottom-up (animal learning theory) approaches to action and decision making are outlined. The specific problems that delays and uncertainty cause for learning and choice are reviewed, with a discussion of the relevance of individual differences in delay and uncertainty processing. The effects of delays and uncertainty upon learning and choice in normal animals are reviewed. Next, systemic psychopharmacological studies examining delayed or uncertain reinforcement are discussed. Neurobiological studies are then examined, beginning with a brief review of the anatomy of relevant parts of the limbic system, followed by an overview of the role of these structures in reinforcement learning in general, and concluding with a review of neuroanatomical studies specifically concerning delayed and/or uncertain reinforcement. An overview of experimental work contained in this thesis is then provided.

1.2 NORMATIVE AND BEHAVIOURAL ECONOMIC APPROACHES TO DECISION MAKING

Behavioural economics is a merging of traditional economic theory with psychological studies of choice (Rachlin *et al.*, 1976; Allison, 1979) that offers a quantitative approach to choice and decision making.

Much of economics is based on utility theory (von Neumann & Morgenstern, 1947; Russell & Norvig, 1995), which assumes that agents are rational in that they exhibit certain reasonable attributes of preference. For example, one assumption is transitivity of preference: if an agent prefers A to B and B to C, then it must prefer A to C (or it would easily be exploited by more rational agents). Given these assumptions, there must exist a utility function that assigns unidimensional values to real-world multidimensional events or outcomes, such that the agent prefers outcomes with higher utility. Psychologically and neurally, a similar process must also happen (Shizgal, 1997)—if at no earlier stage of processing, incompatible behaviours must compete for access to motor output. Agents can then use their knowledge about the world, and about the consequences of their actions (which may be uncertain), to act so as to maximize their expected utility (Arnauld & Nicole, 1662). To allow for the fact that actions may not always have totally predictable consequences, the agent's knowledge about the causal nature of the world may be represented in the form $P(action \rightarrow outcome_n \mid evidence)$ denoting the probability, given the available evidence, that *action* causes $outcome_n$. If $U(outcome_n)$ is the utility of obtaining $outcome_n$, then the expected utility of an action is therefore given by $EU(action \mid evidence) = \sum_n P(action \rightarrow outcome_n \mid evidence) \cdot U(outcome_n)$. Rational decision making follows if the agent selects the action with the maximum expected utility (the MEU principle). The theory specifies neither the utility functions themselves—anything can be valued—nor the way that the decision is arrived at. Rational behaviour need not require complex, explicit thought; merely that observed behaviour follows rational principles.

Conversely, if agents are logical, then we can infer their value system (utility function) by observing their behaviour—the principle of revealed preference (Friedman, 1990; Williams, 1994). To do so, we must assume that agents have reasonably simple objectives (for if we allow that the behaviour we observe is itself the agent's objective, we could explain any arbitrary behaviour).

Assuming rationality allows us to predict behaviour much better than not assuming rationality, unless we can predict the specific way in which people will be irrational (Friedman, 1990). However, humans do not always choose according to rational norms. Introducing an element of randomness into decision making can be theoretically optimal in some situations (von Neumann & Morgenstern, 1947; Mérö, 1998) and the requirement to make rapid decisions may promote the use of heuristics to approximate rational decision making (Russell & Norvig, 1995). Empirically, humans systematically deviate from the optimum when making decisions (Kahneman *et al.*, 1982; Heckerman *et al.*, 1992; Lopes, 1994; Chase *et al.*, 1998; Mullainathan, 2002). They do so because human cognitive abilities are limited ("bounded rationality") and because people frequently make choices that aren't in their long-term interest ("bounded willpower"). In particular, humans and animals do not discount the future in a self-consistent way (Ainslie, 1975; 2001). This point will be expanded upon later, but it serves to illustrate the departure of animal decision making from economic optimality in some situations.

1.3 BASIC PSYCHOLOGY OF REINFORCEMENT LEARNING

An alternative perspective on actions and their consequences stems from the animal learning theory literature. Whereas behavioural economists tend to adopt a top-down approach, treating the agent as a single decision-making entity, animal learning theorists have sought to identify subcomponents and mechanisms giving rise to overt behaviour. The study of motivated action is the study of instrumental conditioning—the process by which animals alter their behaviour when there is a contingency between their behaviour and a reinforcing outcome (Thorndike, 1911). Reinforcement learning (Minsky, 1961; Russell & Norvig, 1995; Haykin, 1999) has been studied for a long time (Thorndike, 1905; Thorndike, 1911; Grindley,

1932; Guthrie, 1935; Skinner, 1938; Hull, 1943). At its most basic level, it is the ability to learn to act on the basis of important outcomes such as reward and punishment; events that strengthen (increase the likelihood of) preceding responses are called positive reinforcers, and events whose removal strengthens preceding responses are called negative reinforcers (Skinner, 1938; 1953). If reinforcers are defined by their effect on behaviour, then, to avoid a circular argument, behaviour cannot be said to have altered as a consequence of reinforcement (Skinner, 1953). However, to explain behaviour, rather than merely to describe it, internal processes such as motivation must also be accounted for. Central motivational states, such as hunger and thirst, are intervening variables that parsimoniously account for a great deal of behavioural variability (Erwin & Ferguson, 1979; Toates, 1986; Ferguson, 2000). For example, water deprivation, eating dry food, hypertonic saline injection, and the hormone angiotensin II all induce a common state (thirst) that has multiple effects: thirsty animals drink more water, drink water faster, perform more of an arbitrary response to gain water, and so on. The ideas of motivational state entered early theories of reinforcement. For example, it was suggested that events that reduce "drive" states such as thirst are positively reinforcing (Hull, 1943). However, on its own this simple model cannot account for many instrumental conditioning phenomena, let alone "unnatural" reinforcement such as intracranial self-stimulation (ICSS) and drug addiction.

Modern neuropsychological theories of instrumental conditioning recognize that many processes contribute to a simple act such as pressing a lever to receive food (e.g. Dickinson, 1994). I will merely summarize these processes here; for a full review, see Cardinal *et al.* (2002a). Rats and humans exhibit goal-directed action, which is based on knowledge of the contingency between one's actions and their outcomes, and knowledge of the value of those outcomes. These two knowledge representations interact so that we work for that which we value (Dickinson, 1994; Dickinson & Balleine, 1994). Environmental stimuli ("discriminative stimuli" or S^Ds) provide information about what contingencies may be in force in a given environment (Colwill & Rescorla, 1990; Rescorla, 1990a; 1990b). Remarkably, the value system governing goal-directed action is not the brain's only one. This "cognitive" value system (sometimes termed "instrumental incentive value") may be distinguished and dissociated (Balleine & Dickinson, 1991) from a different valuation process that determines our reactions when we actually experience a goal such as food—termed "liking", "hedonic reactions", or simply "pleasure" (Garcia, 1989). Under many normal circumstances the two values reflect each other and change together. However, the fact that they are different means that animals must *learn* what outcomes are valuable (hedonically pleasant) in a given motivational state, a process referred to as incentive learning. For example, rats do not know that to eat a particular food while sated is not as valuable as to eat the same food while hungry *until* they have actually eaten the food while sated (Balleine, 1992).

Just as there is more than one value system, there is more than one route to action. Not all action is goal directed. With time and training, actions can become habitual (Adams, 1982)—that is, elicited in relevant situations by direct stimulus–response (S–R) associations. S–R habits are less flexible than goal-directed action, because their representation contains no information about what the outcome will be, and therefore cannot alter quickly if the desirability of a particular outcome changes. However, habits may be important to reduce the demands on the cognitive, goal-directed system in familiar settings.

Finally, environmental stimuli have effects beyond eliciting habits and serving as discriminative stimuli. Stimuli that predict reward may become conditioned stimuli (CSs), associated with the reward (unconditioned stimulus, US) through Pavlovian associative learning (Pavlov, 1927). Pavlovian CSs can elicit Pavlovian conditioned responses (CRs), can influence ongoing instrumental behaviour directly (termed Pavlovian–instrumental transfer or PIT), and can serve as the goals of behaviour (termed condi-

tioned reinforcement) (see Estes, 1948; Lovibond, 1983; Dickinson, 1994; Dickinson & Balleine, 1994; Cardinal *et al.*, 2002a).

1.4 DELAYED AND UNCERTAIN REINFORCEMENT: THE PROBLEMS OF LEARNING AND CHOICE

Natural and artificial learning agents must grapple with the problem of selecting actions to achieve the best possible outcome under their value system. However, the outcome of a given action is not always certain and immediate. Outcomes are frequently uncertain: agents do not always obtain that for which they work. Furthermore, when an agent acts to obtain reward or reinforcement, there is often a delay between its action and the ultimate outcome. This applies both to positive reinforcers (things whose presentation increases the likelihood of preceding actions) and negative reinforcers (things whose removal increases the likelihood of preceding actions) (Skinner, 1938), though I will focus on positive, or appetitive, reinforcers, such as food; I will also use the term "reward" for an appetitive positive reinforcer. For optimal performance, therefore, agents must learn and choose on the basis of reward or reinforcement that is uncertain or delayed.

As discussed above, agents may act procedurally, meaning that they act without an representation of the outcome of their actions, but merely on the basis that an action has been reinforced or led to unspecified "good things" before. Direct links between representations of triggering stimuli and particular responses exemplify procedural responding, or stimulus–response (S–R) learning; the S–R links are strengthened in some way as a result of the arrival of reinforcement, but without the nature of that reinforcement being explicitly encoded. Alternatively, or additionally, agents may encode the outcomes of their actions explicitly, and use these explicit (sometimes termed declarative) representations of anticipated actions when choosing what to do. Animals exhibit both stimulus–response (procedural) and truly goal-directed or action–outcome (declarative) responding (Dickinson, 1994; Dickinson & Balleine, 1994; Cardinal *et al.*, 2002a). This complicates the analysis of motivated behaviour in animals, including the analysis of learning with and choosing uncertain and delayed rewards.

In an S–R learning system, it is easy to envisage connectionist mechanisms by which uncertain and delayed reinforcers could drive learning. Suppose an agent experiences its world, causing many different "stimulus units" to become activated, and suppose it acts randomly by activating different "response units". Let us consider the basic case of appetitive, certain, immediate reinforcement. Suppose a hard-wired mechanism exists to detect events of innate importance to the agent, such as food to a hungry animal. Suppose also that this mechanism, upon detecting an important appetitive event, triggers an internal reinforcement signal that acts to strengthen links between currently active units (stimulus units and response units). By strengthening links between units representing the stimuli currently being perceived and the response currently executing, this simple system would reinforce the response, i.e. increase the probability of executing the response again in the same situation. These S–R links do not encode the nature of the food. If the relationship between responses and food is uncertain, i.e. if $0 < P(\mathit{outcome} \mid \mathit{action}) < 1$, then S–R connections will be reinforced on occasions when food is delivered, but not reinforced on occasions when it is not. S–R links would thus develop to reflect the statistical relationships between actions and reward in a particular stimulus environment: more reliable action–outcome contingencies in the environment come to be reflected in stronger S–R links. To extend this to delayed reinforcement, when the time $t(\mathit{action} \rightarrow \mathit{outcome}) > 0$, requires that some representation of recently executed responses remains active until the reinforcing outcome actually arrives, if the correct response is to be reinforced. If the ac-

tion representation decays gradually (or if it persists until a new action is begun, and the probability of remaining in the same "action state" therefore declines with time), then the likelihood of reinforcing the correct response will decline gradually as action–outcome delays increase, and the agent will learn less well as reinforcement is progressively delayed. None of these ideas are new (Thorndike, 1911; Grindley, 1932; Hull, 1932; Guthrie, 1935; Hull, 1943; Spence, 1956; Mowrer, 1960; Revusky & Garcia, 1970; Mackintosh, 1974; Killeen & Fetterman, 1988).

In a goal-directed (action–outcome) learning system, the agent must encode both the action–outcome relationship and the value of the outcome, and these two representations must interact to determine the probability of selecting a given action (Tolman, 1932; Dickinson, 1980; Dickinson, 1994; Dickinson & Balleine, 1994; Cardinal *et al.*, 2002a). Declarative representations are substantially harder to represent using a simple connectionist framework (Holyoak & Spellman, 1993; Shastri & Ajjanagadde, 1993; Sougné, 1998). The problem of detecting and encoding the action–outcome relationship (the consequences of the agent's actions) is itself complex, but the additional issues concerning uncertain and delayed outcomes are much the same as in the S–R case. That is to say, it may be more difficult to learn that an action causes a given outcome if that outcome is inconsistent or delayed. On top of this, even if the agent knows perfectly well that an action produces an outcome with a certain probability and/or a certain delay, the agent may *value* uncertain or delayed rewards less than certain or immediate rewards, reducing the likelihood of its choosing that action. For example, if we ask a man whether he prefers £10 now or £20 next week, we usually assume that he represents the action–outcome contingencies equally (i.e. that he believes that selecting the "£10 now" option is as likely to produce £10 now as selecting the "£20 next week" option is to produce £20 next week) and that his choice simply reflects the relative value to him of the two options. On the other hand, if we train rats to press levers for (say) immediate and delayed reward, we must bear in mind the possibility of inequalities both in the representation of the action–outcome contingency for immediate and delayed reward, and in the values of the two outcomes—not to mention differences in S–R learning that the delays may engender. There are few mechanistic models of explicit (declarative) delay or uncertainty coding applicable to animal learning, although a recent model proposes the encoding of action uncertainty as a way of mediating competition between goal-directed and S–R responding (Daw *et al.*, 2005).

1.5 INDIVIDUAL DIFFERENCES: RISK TAKING AND IMPULSIVITY

Individual differences in responsivity to uncertain or delayed reinforcement are also of considerable interest. When making decisions under conditions of uncertainty, individuals vary as to how much uncertainty or risk they are willing to tolerate. Formally, individuals differ in how much they discount the value of reinforcers as the uncertainty of the reinforcer increases—i.e. as the probability of the reinforcer declines, or the odds against obtaining the reinforcer increase (Ho *et al.*, 1999). Risk taking is one aspect of the personality trait of impulsivity (Daruna & Barnes, 1993; Eysenck, 1993; Evenden, 1999a) and is a feature of a number of psychiatric disorders, including pathological gambling, antisocial personality disorder, and borderline personality disorder (Roy *et al.*, 1989; Coccaro & Siever, 1995; APA, 2000; Holt *et al.*, 2003). The term "risk" implies exposure to the possibility of an aversive consequence (OUP, 1997), which may include the possibility of not obtaining an anticipated reward. In the appetitive domain, risk taking is exemplified by the tendency to choose large rewards that are very uncertain, in preference to smaller, certain rewards. Abnormal risk taking may reflect dysfunction of reinforcement learning systems that mediate the effects of uncertain reward or punishment.

Furthermore, individual variation in the ability to use delayed reinforcement may determine another aspect of impulsivity: an animal able to forgo short-term poor rewards in order to obtain delayed but better rewards may be termed self-controlled, whereas an animal that cannot tolerate delays to reward may be said to exhibit impulsive choice (Ainslie, 1975; Evenden, 1999b; Evenden, 1999a; Ainslie, 2001). Abnormalities in learning from delayed reinforcement may be of considerable clinical significance (Rahman *et al.*, 2001). Impulsivity is part of the syndrome of many psychiatric disorders, including mania, drug addiction, antisocial personality disorder, borderline personality disorder, and ADHD (APA, 2000). Impulsivity is a broad concept that may be divided into preparation impulsivity (failure to take all relevant information into account before making a decision), execution or "motor" impulsivity (termination of a behavioural chain before the goal is reached), and outcome or "choice" impulsivity (choice of a quick but less valuable outcome rather than a later but more valuable outcome). These measures may be pharmacologically dissociated (Evenden, 1999b; Evenden, 1999a). Impulsive choice, one aspect of impulsivity (Evenden, 1999b), may reflect dysfunction of reinforcement learning systems mediating the effects of delayed rewards (Ainslie, 1975; Sagvolden & Sergeant, 1998).

1.6 LEARNING WITH DELAYED REINFORCEMENT IN NORMAL ANIMALS

1.6.1 Basic phenomena

Delays can hamper both Pavlovian and instrumental conditioning (Dickinson, 1980; Mackintosh, 1983; Dickinson, 1994; Gallistel, 1994; Hall, 1994). For example, instrumental conditioning has long been ob-

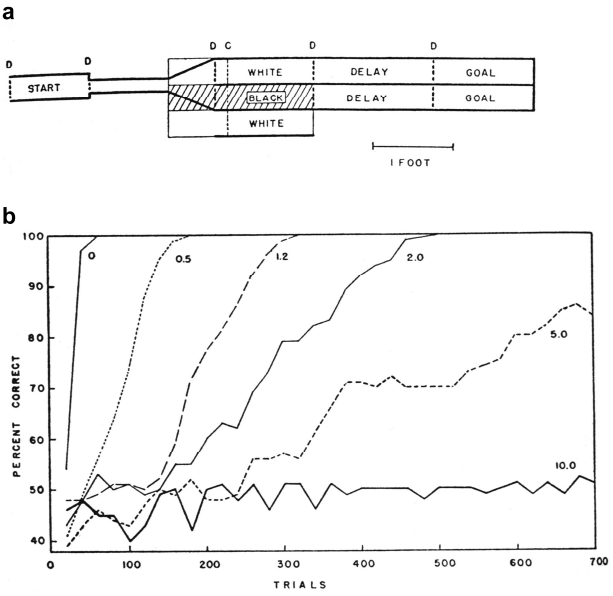

FIG. 2. Learning curves for each of the six different delay groups

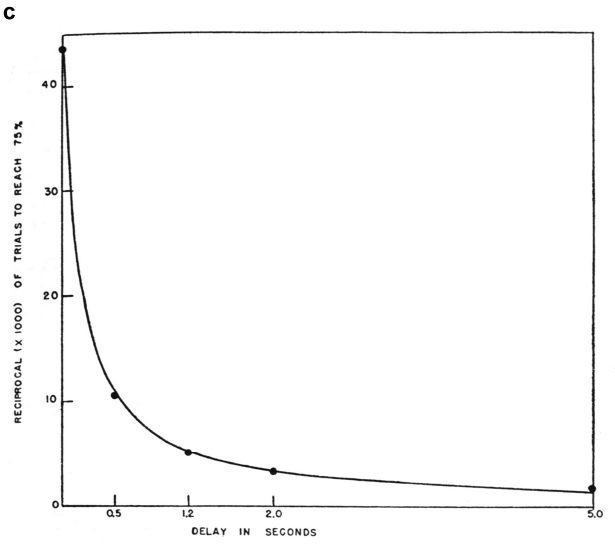

FIG. 3. Rate of learning as a function of delay of reward. The reciprocal × 1000 of the number of trials to reach the level of 75 percent correct choices is plotted against the time of delay. Experimental values are represented by black dots and the smooth curve is fitted to these data.

Figure 1: Discrimination learning with delayed reinforcement

Grice (1948) trained rats on a visual discrimination task with delayed reinforcement. The rats had a choice of a white or a black start alley (which varied in their left/right position); the delay was provided by two grey alleys of variable length which terminated in two grey goal boxes **(a)**. Choosing white led to a goal box with food; choosing black led to an empty box. Grice found that learning was noticeably impaired by as short a delay as 0.5 s, and severely impaired by 5 s, with little learning at a delay of 10 s **(b, c)**. This deficit could be ameliorated by having more discriminable (black and white) goal boxes, or forcing the rats to make discriminable motor responses (climbing an incline or dodging between blocks) in the black and white start alleys (data not shown). Figures from Grice (1948).

served to be systematically impaired as the outcome is delayed (Skinner, 1938; Perin, 1943; Grice, 1948; Harker, 1956; Lattal & Gleeson, 1990; Dickinson *et al.*, 1992) (Figure 1). Despite this, normal rats have been shown to acquire free-operant responding (Figure 2) with programmed response–reinforcer delays of up to 32 s, or even 64 s if the subjects are pre-exposed to the learning environment (Dickinson *et al.*, 1992) (Figure 3). Delays do reduce the asymptotic level of responding (Dickinson *et al.*, 1992), though the reason for this is not clear. There are several psychological reasons why action–outcome delays might impair learning or performance of an instrumental response (Ainslie, 1975; Cardinal *et al.*, 2004). As discussed above, it may be that when subjects learn a response with a substantial response–reinforcer delay, they never succeed in representing the instrumental action–outcome contingency fully. Alternatively, they may value the delayed reinforcer less. Finally, the delay may also retard the acquisition of a procedural stimulus–response habit and this might account for the decrease in asymptotic responding. It is presently not known whether responses acquired with delayed reinforcement are governed by a different balance of habits and goal-directed actions than responses acquired with immediate reinforcement.

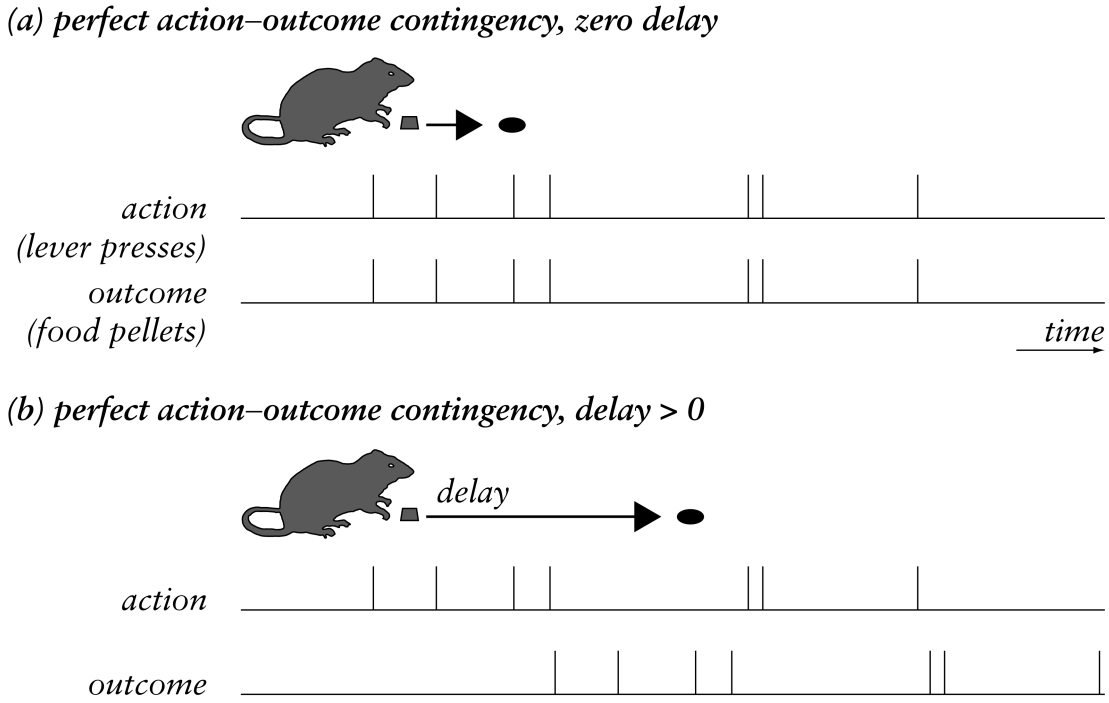

Figure 2: Free-operant learning with delayed reinforcement

When an animal is free to perform an action (operant) to obtain a rewarding outcome, it readily learns to do so if the action–outcome contingency (the increase in the likelihood of obtaining the outcome that is produced by performing the action) is good and if there is no delay between action and outcome **(a)**. Even with a perfect action–outcome contingency, learning is impaired by imposing delays between the action and the outcome **(b)**, yet animals do succeed in this task.

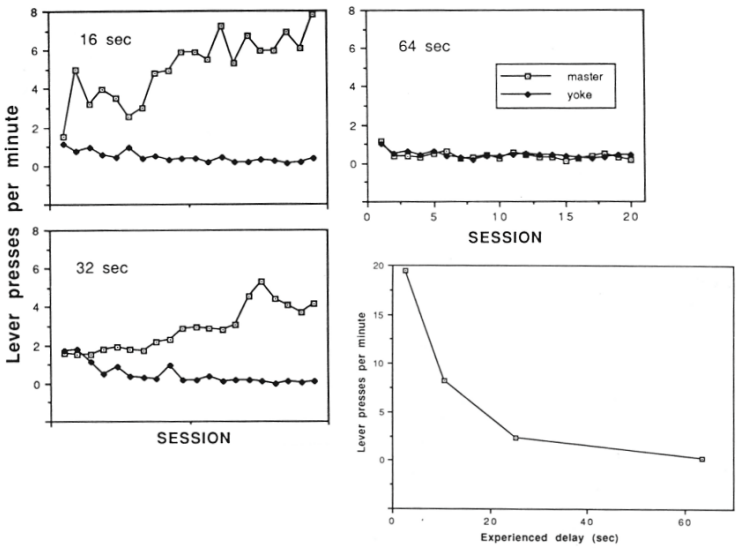

Figure 3: The speed of free-operant learning with delayed reinforcement in normal rats

Dickinson *et al.* (1992) trained rats on a free-operant, fixed-ratio-1 (FR-1) schedule of reinforcement with delays between pressing the lever and obtaining reinforcement (see Figure 2). Responding was compared to that of a yoked control group (who received the same pattern of reinforcement as the "master" rats but whose lever presses had no consequence). The rate of learning, and the asymptotic level of responding, declined across groups as the response–reinforcer delay was increased from 0 to 32 s; rats trained with a 64-s delay failed to learn at all, compared to yoked controls. However, when rats were exposed to the training context, in the absence of the lever or any reinforcers, prior to training, their learning was improved, and successful discrimination was seen even with a delay of 64 s (data not shown), attributed to an underlying process of contextual competition (see text). From Dickinson *et al.* (1992).

1.6.2 Cues and context

Two additional factors must be considered. Cues or signals present during the delay to the reinforcer may become associated with the primary reinforcer, becoming conditioned reinforcers capable of reinforcing actions themselves; conditioned reinforcers may therefore help to bridge action–outcome delays. Indeed, such signals tend to increase responding for delayed reinforcers (Lattal, 1987; Mazur, 1997). One other important factor in learning to act using delayed reinforcement may be the role of the environmental context. The animal's task is to attribute the outcome to its actions; instead, it may erroneously associate the outcome with the context, since the context is a cue that is temporally closer to the outcome than the action is. The longer the delay, the more this contextual competition comes to impair the learning of the action–outcome contingency. Instrumental conditioning with delayed reinforcement can be enhanced if rats are exposed to the relevant contextual cues prior to instrumental training, and this enhancement is lessened if "free" (non-contingent) rewards are given during the contextual pre-exposure periods (Dickinson *et al.*, 1992; Dickinson & Balleine, 1994). These results are consistent with the theory that during the action–outcome delay, contextual cues compete with the action to become associated with the outcome; pre-exposing the animals to the context with no consequences reduces this contextual competition, by making the context a bad predictor of the outcome (perhaps via latent inhibition or learned irrelevance), and this in turn makes the action–outcome contingency more salient and easier to learn (Dickinson *et al.*, 1992; Dickinson & Balleine, 1994).

1.7 CHOICE WITH DELAYED AND UNCERTAIN REINFORCEMENT IN NORMAL ANIMALS

1.7.1 Delayed and probabilistic reinforcement: equivalent or distinct processes?

It has been suggested that delay (or temporal) discounting, the process by which delayed reinforcers lose value, and probability (or odds) discounting, the process by which uncertain reinforcers lose value, reflect the same underlying process (Rotter, 1954; Mischel, 1966; Rachlin *et al.*, 1986; Stevenson, 1986; Rachlin *et al.*, 1987; Mazur, 1989; Rachlin *et al.*, 1991; Mazur, 1995; Green & Myerson, 1996; Mazur, 1997; Sozou, 1998). For example, choosing an uncertain reinforcer five times but only obtaining it on the fifth response might be seen as equivalent to a very long delay, on average, between choice of the reinforcer and its eventual delivery. Alternatively, delays may be seen as entailing the ecological risk of losing the reward during the delay. In animal models, while subjects are learning to respond for delayed or probabilistic rewards, both may initially be similarly unpredictable (although delayed rewards can become more accurately predicted following learning in a manner that stochastic rewards cannot). However, there is evidence that time and probability discounting are different and dissociable processes (Ho *et al.*, 1999; Mitchell, 2003; Green & Myerson, 2004). Most simply, it is not surprising that currency inflation affects human decisions involving delayed but not probabilistic financial reward (Ostaszewski *et al.*, 1998). Moreover, the absolute magnitude of rewards can have different effects on delayed and probabilistic discounting (Green *et al.*, 1999; Myerson *et al.*, 2003; Green & Myerson, 2004). A study looking at human choices in a gambling task found that individuals' propensity to choose rapidly (one, perhaps motoric, measure of delay aversion) and their propensity to bet large amounts of money on uncertain outcomes (a measure of risk taking) represented independent factors (Deakin *et al.*, 2004). Some studies have found abnormal delay discounting, but not uncertainty discounting, in drug addicts (Vuchinich & Calamas, 1997; Mitchell, 1999; Mitchell, 2003; Reynolds *et al.*, 2004b), while gamblers have been observed to discount probabilistic rewards less steeply than controls (i.e. to take risks) without showing differences in delay discounting (Holt *et al.*, 2003).

1.7.2 Temporal or delay discounting

In a typical delayed reinforcement choice task, a subject chooses between an immediate, small ("smaller, sooner" or SS) reward or a large, delayed ("larger, later" or LL) reward; the temporal discounting function quantifies the effect of the delay on preference (Figure 4). Early models of choice assumed an exponential model of temporal discounting (see Kacelnik, 1997a), so that if V_0 is the value of a reinforcer delivered immediately, then the value of a reinforcer delivered after time t is

$$V_t = V_0 e^{-Kt}$$

where K quantifies an individual's tendency to "discount" the future (to value delayed rewards less). The exponential model makes intuitive sense, whether you consider the underlying process to be one in which the subject has a constant probability of "forgetting" its original response per unit time (making it progressively less available for reinforcement), one in which the "strength" of the response's representation decays to a certain proportion of its previous value at each time step, or one in which the subject behaves as if there is a constant probability of losing the delayed reward per unit of waiting time. A S–R learning 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

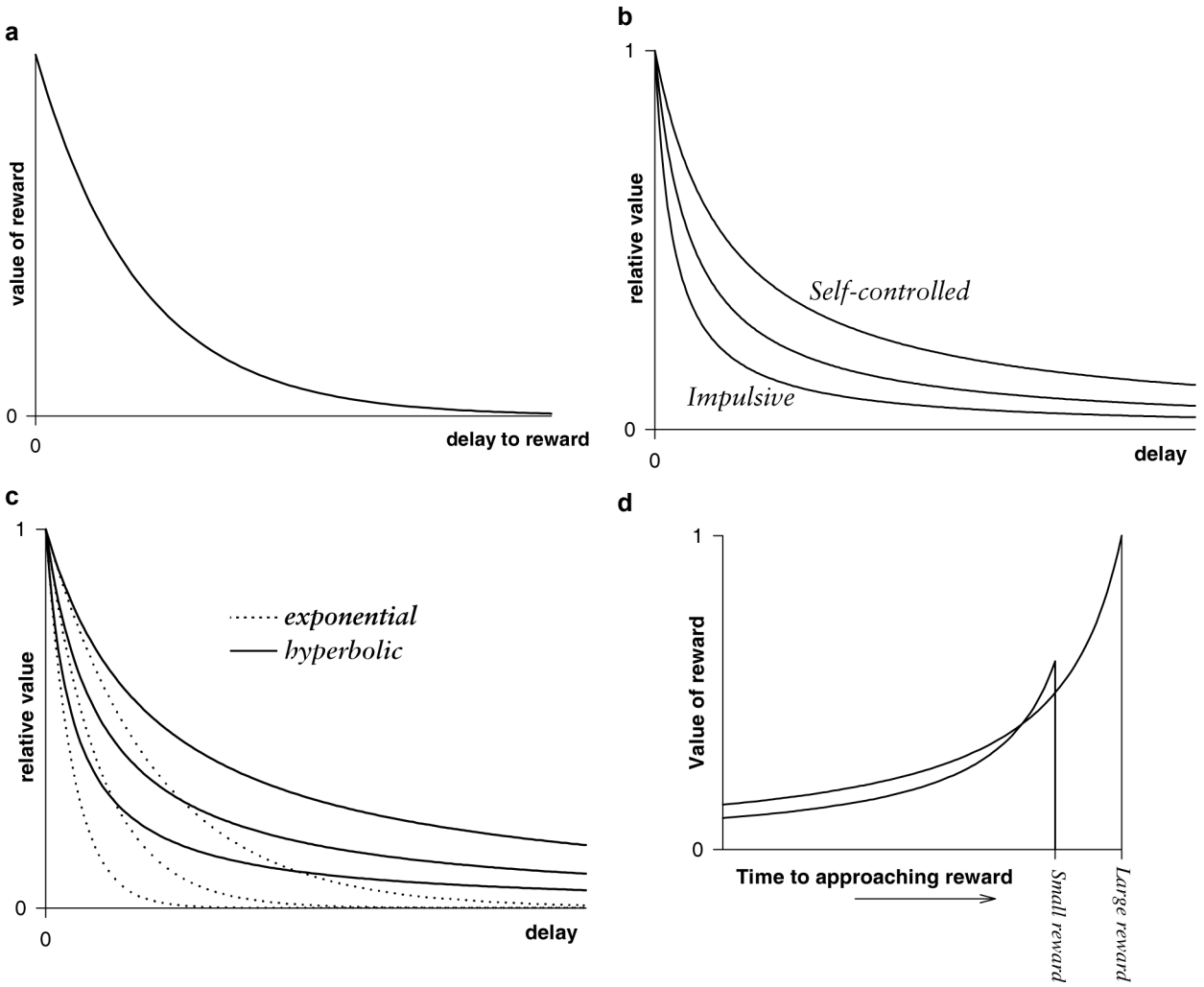

Figure 4: Temporal discounting

(a) The basic, intuitive, and well-validated phenomenon of temporal discounting is that the subjective value of a reward declines monotonically as the reward is progressively delayed: all other things being equal, immediate rewards are worth more than delayed rewards. **(b)** Individuals may vary in their propensity to discount delayed rewards. Individuals who discount the future steeply are said to be impulsive; individuals who discount the future shallowly (giving the future relatively greater weight) are said to be self-controlled. **(c)** Different mathematical models of temporal discounting have been proposed; exponential and hyperbolic discounting are shown. Exponential temporal discounting is described by the equation *value* = *immediate value* × $e^{-K \cdot delay}$. Hyperbolic temporal discounting is governed by the equation *value* = *immediate value* / (1 + $K \cdot delay$). Large values of the discounting parameter K give the steepest curve (the most impulsive behaviour) in both cases. There is strong empirical support for the hyperbolic, not the exponential, discounting model. One critical difference in the predictions of these two models is the phenomenon of preference reversal, since hyperbolic discounting allows curves for different rewards to cross. **(d)** Preference reversal, illustrated for two hypothetical rewards. Given a choice between an early reward of value 0.6 and a later reward of value 1, hyperbolic discounting predicts that the larger reward will be chosen if the choice is made far in advance (towards the left of the graph). However, as time advances, there may come a time just before delivery of the small reward when the value of the small reward exceeds that of the large reward; preference reverses and the small reward is chosen. Figures adapted from Ainslie (1975) (and also published in Robbins *et al.*, 2005).

strength at the previous instant. If x_t is the trace strength at time t and A is the starting value, then

$$x_t = x_0 e^{-kt}$$

and

$$x_{t+1} = e^{-k} x_t$$

However, the exponential model has been emphatically rejected by experimental work with humans and other animals. Instead, temporal discounting appears to follow a hyperbolic or very similar discount function, such as

$$V = \frac{V_0}{1 + Kt}$$

(Grice, 1948; Mazur, 1987; Mazur *et al.*, 1987; Grace, 1996; Richards *et al.*, 1997b).

One interesting prediction that emerges from hyperbolic (but not exponential) models is that preference between a large and a small reward should be observed to reverse depending on the time that the choice is made (Figure 4d), and such preference reversal is a reliable and important experimental finding (see Ainslie, 1975; Green *et al.*, 1981; Bradshaw & Szabadi, 1992; Ainslie, 2001). For example, humans generally prefer £100 now to £200 in three years' time, but also generally prefer £200 in nine years to £100 in six years, despite this being the same choice viewed at a different time. If you are aware that your preference may change in this way, you may be able to improve your happiness in the long run by using self-control strategies (Ainslie & Monterosso, 2003). Ainslie (2001) refers to this as bargaining with your future self; the most famous example of precommitment is that of Odysseus (Homer, ~800 BC / 1996, book 12, translation lines 44–60 and 172–217) (Figure 5); others include the use of disulfiram by alcoholics and social precommitment by announcing publicly one's intention to diet. Even pigeons have been observed to use the self-control strategy of precommitment. When pigeons choose between SS and LL rewards, they are often impulsive (Rachlin & Green, 1972), choosing the SS reward, but they have been shown to work to avoid being offered the option of choosing the SS alternative (Ainslie, 1974; Ainslie & Herrnstein, 1981).

It is also worth noting that in the hyperbolic discounting model and all others in which preference reversal occurs, the value at any one moment cannot be calculated directly from the value immediately preceding it in time; therefore, hyperbolic discounting implies that more information is being maintained by the agent than is required for exponential discounting.

It is not known why hyperbolic discounting arises (Kacelnik, 1997a), or what neuropsychological processes are responsible for it. Such discounting might in principle result from some combination of poor knowledge of the contingencies between actions and their outcomes at long delays, weak S–R habits, or because subjects are perfectly aware that the delayed reward is available but assign a low value to it (Cardinal *et al.*, 2003b). Hyperbolic discounting might also be explicable as the overall effect of two or more different systems—for example, a cognitive, declarative system that exhibits minimal or exponential discounting, plus phenomena such as PIT, conditioned salience, or "visceral factors" that make rewards more salient and promote their choice when they are immediately available (Loewenstein, 1996; Cardinal *et al.*, 2003b; Gjelsvik, 2003; Loewenstein & O'Donoghue, 2004). As discussed above, perhaps the most obvious difference between studies of human impulsive choice and animal models is that humans can be offered explicit choices (hypothetical or real: the difference does not appear to be important; Lagorio & Madden, 2005) without prior experience of the situation (Rachlin *et al.*, 1991; Myerson & Green, 1995; de Wit *et al.*, 2002)—"pre-packaged" action–outcome contingencies. Other animals must learn these contingencies through experience, implying that the whole gamut of psychological representations that contribute to their actions (including goal-directed actions, S–R habits, and conditioned reinforcers) can influence their choices. Nevertheless, hyperbolic discounting has been observed in humans and other experimental animals.

a

b

Figure 5: An early example of precommitment

(a) Waterhouse (1891), *Ulysses and the Sirens*, depicting Odysseus (Ulysses) from Homer's (~800 BC / 1996) *Odyssey* (book 12, translation lines 44–60 and 172–217). Aware that the Sirens—originally bird-like creatures in Greek mythology—would lure his ship onto the rocks through the eldritch influence of their song on men's minds, yet wishing to hear their song for himself, Odysseus commands his men to stop up their ears and lash him to the mast. He gives them strict instructions not to untie him until they are safely past the sirens, and to ignore any further instructions from him until that point. **(b)** Later painters depicted the sirens in more human fashion, or as mermaids; this is Draper's (1909) painting of the same title.

1.7.3 Uncertainty discounting

Similarly, the dominant model of uncertainty or probability discounting (Rachlin *et al.*, 1986; Rachlin *et al.*, 1991; Ho *et al.*, 1999; Green & Myerson, 2004) suggests that subjects calculate a value for each reinforcer, according to its size and other parameters, and discount this by multiplying it by $1/(1+H\theta)$. In this equation, θ represents the odds against obtaining the reinforcer, $\theta = (1 - p)/p$, where p is the probability of obtaining the reward, and H represents an odds discounting parameter that is specific to the individual subject but stable over time for that subject. In this model, value is a hyperbolic function of the odds θ:

$$V = \frac{magnitude}{1 + H \cdot \theta}$$

Such a hyperbolic function is supported by empirical research, at least in humans (Rachlin *et al.*, 1986; Rachlin *et al.*, 1991; Rachlin & Siegel, 1994; Kacelnik, 1997b; Richards *et al.*, 1999b; Rachlin *et al.*, 2000). Preference reversal effects are also observed in choice under risk or uncertainty (Slovic & Lichtenstein, 1983; Lopes, 1994), with subjects preferring gambles with a low probability of winning a large prize when asked to assign monetary values to the gambles, but then preferring gambles with moderate probabilities and prizes when faced with a direct choice—that is, the task used to measure preference alters that preference. Ho *et al*. (1999) suggested that hyperbolic processes of discounting apply to the delay, probability (odds), and magnitude of a reward, and that these three discounting processes are independent, multiplicative, and each governed by its own discounting parameter (K for delay, H for probability/odds, Q for magnitude) that is relatively stable for an individual. Their combined model is therefore as follows:

$$V = \frac{1}{1 + K \cdot delay} \times \frac{1}{1 + H \cdot \theta} \times \frac{magnitude}{magnitude + Q}$$

It should be noted in passing that although effects of delay, probability, magnitude, and so forth are often assumed to be calculated independently (Killeen, 1972; Rachlin *et al.*, 1991; Ho *et al.*, 1999), and though there is some support for this assumption (Mazur, 1987; Mazur, 1997), others have found that the effects of reinforcer delay and magnitude are not independent (Ito, 1985; White & Pipe, 1987). In addition, as discussed below in the context of drug addiction, humans may show quantitatively different temporal (delay) discounting for qualitatively different reinforcers, such as drugs and money. Furthermore, deprivation of one commodity can selectively increase preference for SS over LL rewards for that commodity (e.g. Mitchell, 2004a), suggesting that parameters such as K and/or Q are not unitary parameters that apply to all reinforcers, and/or that additional parameters specific to reinforcer classes must be added to characterize behaviour fully.

1.8 SYSTEMIC PHARMACOLOGICAL STUDIES OF DELAYED OR UNCERTAIN REINFORCEMENT

Given the importance of impulsive choice in disorders such as addiction (Poulos *et al.*, 1995; Heyman, 1996; Bickel *et al.*, 1999; Evenden, 1999a; Mitchell, 1999) and ADHD (Sagvolden *et al.*, 1998; Sagvolden & Sergeant, 1998), a number of groups have studied the effects on impulsive choice of manipulating neurochemical and neuroanatomical systems implicated in these disorders. I will review pharmacological and neurochemical studies first. To date, more have examined choice involving delayed reinforcement than choice involving uncertain reinforcement, and many more have used appetitive positive reinforcement (reward) than aversive reinforcement such as punishment.

1.8.1 Serotonin (5-HT)

Serotonin (5-hydroxytryptamine, 5-HT) has long been implicated in impulse control. Drugs that suppress 5-HT function were observed to reduce behavioural inhibition, making animals more impulsive in a "motor" sense, as defined above (Soubrié, 1986; Evenden, 1999b). Correlational studies have indicated that low cerebrospinal fluid (CSF) levels of the 5-HT metabolite 5-hydroxyindoleacetic acid (5-HIAA)

are associated with risk taking in monkeys (Mehlman *et al.*, 1994; Evenden, 1998) and impulsive aggression, violence, and suicide in humans (Åsberg *et al.*, 1976; Linnoila *et al.*, 1983; Brown & Linnoila, 1990; Linnoila *et al.*, 1993; Mann, 2003).

Forebrain 5-HT depletion leads to impulsive choice in a variety of paradigms (Wogar *et al.*, 1993; Richards & Seiden, 1995; Bizot *et al.*, 1999; Mobini *et al.*, 2000b) and has been suggested to steepen the temporal discounting function, such that delayed rewards lose their capacity to motivate or reinforce behaviour (Wogar *et al.*, 1993; Ho *et al.*, 1999; Mobini *et al.*, 2000a). The 5-HT-depleted animal becomes hypersensitive to delays, or hyposensitive to delayed reward. As delayed rewards have unusually low value, the animal chooses SS rewards over LL rewards, a characteristic of impulsivity (Ainslie, 1975). Conversely, increasing 5-HT function with the 5-HT indirect agonist fenfluramine decreases impulsive choice (Poulos *et al.*, 1996). Since choice between SS and LL rewards may be affected by changes in the sensitivity to reinforcer magnitude as well as reinforcer delay (Ho *et al.*, 1999), it is important to note that 5-HT depletion does not appear to alter reinforcer magnitude discrimination (Mobini *et al.*, 2000a; Mobini *et al.*, 2000b).

Altered 5-HT function has also been strongly implicated in depression (see e.g. Delgado *et al.*, 1990; Feldman *et al.*, 1997, pp. 842–847; Caspi *et al.*, 2003), but the relationship between depression, impulsivity, and 5-HT is complex. The precise neurochemical abnormality or set of abnormalities in depression is far from clear (e.g. Feldman *et al.*, 1997; Dhaenen, 2001; Stockmeier, 2003). There is no clear-cut relationship between depression itself and levels of 5-HIAA in the CSF (Åsberg, 1997; Feldman *et al.*, 1997, p. 843), although antidepressant drugs themselves tend to lower CSF 5-HIAA (see Bäckman *et al.*, 2000). However, there is a consistent association between low CSF 5-HIAA and suicidal behaviour—not only in depression, but also in schizophrenia and other disorders (see Träskman-Bendz *et al.*, 1986; Cooper *et al.*, 1992; Åsberg, 1997; Cremniter *et al.*, 1999). Patients who are prone to suicide, many of whom are depressed, show high impulsivity (Plutchik & Van Praag, 1989; Apter *et al.*, 1993; Corruble *et al.*, 2003). Thus, low 5-HT function has been linked with impulsive behaviour, which is a risk factor for suicide, and abnormalities of the 5-HT system are also associated with depression, also a strong risk factor for suicide.

However, the results relating 5-HT to impulsivity are not wholly clear-cut. The effects of forebrain 5-HT depletion to promote impulsive choice have sometimes been transient (Bizot *et al.*, 1999) or not observed (Winstanley *et al.*, 2003), and a nonselective 5-HT antagonist has been observed to promote self-controlled choice (Evenden & Ryan, 1996). In humans, lowing 5-HT levels via dietary tryptophan depletion (Biggio *et al.*, 1974; Clemens *et al.*, 1980; Delgado *et al.*, 1989) decreases levels of 5-HT metabolites in CSF (Carpenter *et al.*, 1998; Williams *et al.*, 1999), an indirect indicator of brain 5-HT levels. However, although tryptophan depletion may increase "motor" impulsivity in some tasks (Walderhaug *et al.*, 2002), it does not affect stop-signal reaction time (Clark *et al.*, 2005; Cools *et al.*, 2005), a basic measure of motor control, and it has not been shown to increase impulsive choice in humans (Crean *et al.*, 2002). Likewise, a recent rodent study found that forebrain 5-HT depletion increased motor impulsivity but not delay discounting (Winstanley *et al.*, 2004a). Furthermore, 5-HT efflux in prefrontal cortex (PFC), as measured by microdialysis (as opposed to CSF metabolite levels or whole-tissue post mortem measurement) centred on the prelimbic cortex (PrL), was unexpectedly found to be positively correlated with premature responding in an attentional task, a form of motor impulsivity (Dalley *et al.*, 2002). Post mortem analysis of the same subjects failed to show differences in total tissue 5-HT or 5-HIAA levels between the more impulsive and more self-controlled subgroups. 5-HT may modulate impulsivity in different ways depending on the involvement of different receptor subtypes (Evenden, 1999b; Evenden & Ryan, 1999; Winstanley *et al.*, 2004c). Furthermore, the acute effects of serotonergic drugs on impulsivity

can be the opposite of the chronic effects (Liu *et al.*, 2004), with evidence for complex adaptations within the PFC 5-HT system.

Although manipulations of 5-HT have influenced choice involving delayed reinforcement, there is less evidence that they influence choice involving uncertainty and risk. Although forebrain 5-HT depletion has affected temporal (delay) discounting, as discussed above, it does not appear to influence choice involving probabilistic reinforcement. Dietary tryptophan depletion has not been shown to affect probability discounting in humans (Anderson *et al.*, 2003; Rogers *et al.*, 2003; but see Cools *et al.*, 2005); similarly, forebrain 5-HT depletion in rats does not affect choice between small, certain rewards and large, uncertain rewards (Mobini *et al.*, 2000b).

1.8.2 Noradrenaline (NA)

Relatively little is known about the role of noradrenaline (NA) in delayed or probabilistic reinforcement. It has been suggested that NA neurons encode some aspects of uncertainty in the general sense of making predictions in a given context, in a manner complementary to that of acetylcholine (ACh) (Yu & Dayan, 2005). In causal studies, systemic NA blockade has been shown to affect decision making under uncertainty in humans, by reducing the discrimination between magnitudes of different losses when the probability of losing was high (Rogers *et al.*, 2004a), though NA reuptake inhibition has not been shown to affect the Iowa gambling task (O'Carroll & Papps, 2003), in which subjects must choose between decks of cards differing in magnitude and probability of their expected gains and losses (Bechara *et al.*, 1994).

1.8.3 Dopamine (DA)

1.8.3.1 Temporal difference learning and dopamine

Since prediction of the future is of key importance in designing artificial intelligence agents, a number of mathematical and computational models have been developed to learn from delayed and/or probabilistic reinforcement (Russell & Norvig, 1995), including some forms of Q-learning (Watkins, 1989) and temporal difference (TD) learning (Sutton, 1988). Some models have been compared directly to mammalian neural systems. For example, the TD learning model of Sutton (1988) has been extended to an actor–critic architecture (see Barto, 1995; Houk *et al.*, 1995). In this scheme, a "critic" has access to sensory and motor information and primary reinforcement, and learns to predict reward on the basis of this information using a TD algorithm. "Immediate" reinforcement is held to follow the causing action by one time unit, and the reinforcement at time t is referred to as r_t. Delayed reinforcement is given a lesser weighting by being multiplied by a factor γ for every time step it is delayed (where $0 \leq \gamma < 1$); high γ indicates a strategic or long-term orientation and low γ indicates a tactical, short-term, or impulsive orientation. If the critic is perfect, then its prediction P would be

$$P_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \ldots$$

Therefore, the prediction for time t–1 would be

$$P_{t-1} = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots$$

and thus, for perfect prediction,

$$P_{t-1} = r_t + \gamma(r_{t+1} + \gamma r_{t+2} + \ldots)$$

$$P_{t-1} = r_t + \gamma P_t$$

$$r_t + \gamma P_t - P_{t-1} = 0$$

The TD error δ can therefore be defined as

$$\delta = r_t + \gamma P_t - P_{t-1}$$

This quantity δ represents the difference between predicted and actual reward. The critic learns by adjusting its reinforcement prediction on the basis of the TD error: if $\delta > 0$, reward occurred that was not predicted, and the prediction at t–1 should be increased for next time; if $\delta < 0$, reward was predicted but did not occur, and the prediction at t–1 should be decreased. The critic teaches not only itself but also an "actor", which selects an action, and then modifies the propensity to perform that action on the basis of the TD error (if $\delta = 0$, the consequences of the last action were expected; if $\delta > 0$, the consequences were better than expected, and the response tendency of the action made at t–1 should be strengthened; if $\delta < 0$, the consequences were worse than expected, and the response tendency at t–1 should be decreased).

The result is that if a consistent sequence of stimuli predicts reward, this system will learn the sequence, with the TD error teaching the system about earlier and earlier consistent predictors with each iteration. As the critic learns about future rewards, it is able to teach the actor to act on the basis of them. Thus the system exemplifies S–R learning with an enhanced ability to act on the basis of future reward. It has been of particular neurobiological interest since the firing of midbrain dopamine (DA) neurons appears to correspond very closely to the TD error δ (Schultz *et al.*, 1997; Schultz, 1998; Schultz *et al.*, 1998; Schultz & Dickinson, 2000; Daw & Touretzky, 2002; McClure *et al.*, 2003b; Schultz, 2006), and other components of the basal ganglia innervated by midbrain DA neurons have been proposed to correspond to the actor and critic, be those components the matrix and striosome compartments of the striatum (Houk *et al.*, 1995) or the dorsal and ventral striatum (O'Doherty *et al.*, 2004).

1.8.3.2 Psychostimulants and impulsivity

However, the original interest in the relationship between DA and impulsivity stems from the discovery that amphetamine and similar psychostimulants are an effective therapy for ADHD (Bradley, 1937). Though these drugs have many actions, they are powerful releasers of DA from storage vesicles in the terminals of dopaminergic neurons, and prevent DA re-uptake from the synaptic cleft, potentiating its action (for references see Feldman *et al.*, 1997, pp. 293/552/558). It has been proposed that many features of ADHD, including preference for immediate reinforcement and hyperactivity on simple reinforcement schedules, are due to abnormally steep temporal discounting, and that this is due to a hypofunctional nucleus accumbens (Acb) DA system (Sagvolden *et al.*, 1998; Sagvolden & Sergeant, 1998; Johansen *et al.*, 2002). Indeed, they go on to suggest Acb DA as the specific culprit (Sagvolden *et al.*, 1998; Sagvolden & Sergeant, 1998). Acb DA has long been implicated in aspects of responding for reinforcement, though its role is not yet fully understood (Cardinal *et al.*, 2002a; Salamone *et al.*, 2005). However, whether ADHD is characterized by a hypodopaminergic or a hyperdopaminergic state, and how this and other (e.g. noradrenergic/serotonergic) abnormalities might be "normalized" by psychostimulants is controversial (Swanson *et al.*, 1998; Zhuang *et al.*, 2001; Seeman & Madras, 2002; Solanto, 2002; Fone & Nutt, 2005; Russell *et al.*, 2005; Williams & Dayan, 2005).

Many of the inferences regarding the neural abnormalities in children with ADHD have been drawn from studies of the spontaneously hypertensive rat (SHR), an inbred strain of rat that serves as an animal

model of ADHD (Wultz *et al.*, 1990; Sagvolden *et al.*, 1992; Sagvolden *et al.*, 1993; Sagvolden, 2000; Russell *et al.*, 2005). This rat exhibits pervasive hyperactivity and attention problems that resemble ADHD, exhibits a steeper "scallop" of responding on fixed-interval (FI) schedules of reinforcement, which can be interpreted as abnormally high sensitivity to immediate reinforcement (Sagvolden *et al.*, 1992), is impulsive on measures of "execution impulsivity" (Evenden & Meyerson, 1999), and has a complex pattern of abnormalities in its DA system (de Villiers *et al.*, 1995; Russell *et al.*, 1995; Papa *et al.*, 1996; Carey *et al.*, 1998; Papa *et al.*, 1998; Russell *et al.*, 1998; Russell, 2000). Depolarization- and psychostimulant-induced DA release in Acb brain slices is altered in the SHR compared to Wistar Kyoto progenitor control rats in a complex pattern that has been attributed to hypofunction of the mesolimbic DA system projecting to the Acb (de Villiers *et al.*, 1995; Russell *et al.*, 1998; Russell, 2000), though abnormalities have also been found in DA release in slices of dorsal striatum and PFC (Russell *et al.*, 1995). Within the Acb, differences in gene expression and DA receptor density have been observed in both the core and shell subregions (Papa *et al.*, 1996; Carey *et al.*, 1998; Papa *et al.*, 1998).

Impulsive choice may reflect a lack of effectiveness of delayed reinforcement, and has been suggested to underlie at least some subtypes of ADHD (Sagvolden *et al.*, 1998; Sagvolden & Sergeant, 1998; Kuntsi *et al.*, 2001; Sonuga-Barke, 2002). The efficacy of psychomotor stimulants in ADHD (Bradley, 1937; Solanto, 1998) suggests that they might promote the choice of delayed rewards. In fact, the effects of acute administration of psychostimulants on laboratory models of impulsive choice have varied. Some studies have found that they promote choice of delayed reinforcers (Sagvolden *et al.*, 1992; Richards *et al.*, 1997a; Richards *et al.*, 1999a; Wade *et al.*, 2000; de Wit *et al.*, 2002), while others have found the opposite effect (Logue *et al.*, 1992; Charrier & Thiébot, 1996; Evenden & Ryan, 1996); the same psychostimulant can even have opposite effects in different tasks designed to measure impulsivity (Richards *et al.*, 1997a). One factor that may explain some of these discrepant effects is the presence of cues or signals present during the delay to the larger/later alternative. Such signals tend to increase responding for the delayed reinforcer (Lattal, 1987; Mazur, 1997), perhaps because they become associated with the primary reinforcer and themselves become conditioned reinforcers, thus affecting choice (Williams & Dunn, 1991). Psychostimulants increase the effect of conditioned reinforcers (Hill, 1970; Robbins, 1976; Robbins, 1978; Robbins *et al.*, 1983), and their effects in delayed reinforcement choice tasks can depend on whether explicit signals are presented during the delay (Cardinal *et al.*, 2000). However, conditioned reinforcement is certainly not the only procedural difference between studies that have found differing effects of psychostimulants.

1.8.3.3 Dopamine D_1 and D_2 receptors and impulsivity

It should also be emphasized that few studies of the effects of psychostimulants on impulsive choice have addressed the pharmacological basis of their effects. Some of the effects may indeed not be dopaminergic: for example, the effects of amphetamine may depend in part on 5-HT (Winstanley *et al.*, 2003). However, Wade *et al*. (2000) have shown that D_2-type DA receptor antagonists and mixed D_1/D_2 antagonists induce impulsive choice, while D_1-type receptor antagonists do not, suggesting that DA D_2 receptors normally promote choice of delayed reinforcement.

The role of DA in reward uncertainty is also not well understood. DA neurons respond to reward prediction errors with changes in their phasic firing rate, as discussed above, and may also carry information in their sustained firing rate specifically about reward uncertainty (Fiorillo *et al.*, 2003; Fiorillo *et al.*, 2005; Niv *et al.*, 2005; Tobler *et al.*, 2005), but little is known of the causal role of DA in choice involving uncertain rewards.

1.8.4 Relationship between addictive drugs and impulsivity

Given that impulsivity is part of the syndrome of drug addiction, with impulsive choice playing a prominent role in maintaining the selection of drugs of abuse in favour of other, longer-term rewards (Poulos *et al.*, 1995; Heyman, 1996; Bickel *et al.*, 1999; Evenden, 1999a; Mitchell, 1999; APA, 2000), the relationship between addictive drugs and impulsive choice is of clear interest. Studies examining discounting in addicts have focused primarily on delay, rather than uncertainty discounting (see Mitchell, 2003; 2004b; 2004a). There is little evidence for differences in uncertainty discounting among smokers (Mitchell, 1999) or alcohol abusers (Vuchinich & Calamas, 1997), though alcohol has been shown to modify decision making under uncertainty (George *et al.*, 2005). However, decision-making deficits in risk-taking and gambling tasks have been demonstrated in opiate and amphetamine users (Rogers *et al.*, 1999a; Ersche *et al.*, 2005; Leland & Paulus, 2005). The fact that the deficits were in some cases correlated with the number of years of abuse suggests (but does not prove) that the deficits were drug-induced; the possibility remains that the decision-making deficits predated and predisposed to the addiction.

Abnormally steep delay discounting has been demonstrated in drug addicts, including alcoholics (Vuchinich & Calamas, 1997; Petry, 2001), cocaine users (Coffey *et al.*, 2003; Kirby & Petry, 2004), opiate users (Madden *et al.*, 1997; Kirby *et al.*, 1999; Kirby & Petry, 2004), and smokers (Bickel *et al.*, 1999; Mitchell, 1999; Mitchell, 2003; Reynolds *et al.*, 2004b; Ohmura *et al.*, 2005; Reynolds, 2006); again, the question of cause and effect is hard to determine, although steeper discounting in current addicts compared to ex-addicts again raises the possibility of an effect of ongoing drug use. Many studies have looked at the pharmacological effects of addictive drugs on measures of impulsivity including response inhibition; rather fewer have looked specifically at delay and/or probability discounting in a formal experimental (causal) design. Chronic cocaine administration transiently increases delay discounting (increases impulsive choice) in rats (Paine *et al.*, 2003), as does acute morphine administration (Kieres *et al.*, 2004); acute administration of psychostimulants was discussed above, and chronic methamphetamine has been shown to increase impulsive choice in rats (Richards *et al.*, 1999a). In keeping with everyday experience, alcohol has been observed to increase risk taking (Lane *et al.*, 2004). However, the findings for a given drug have not always been consistent. For example, Ortner *et al*. (2003) recently found that alcohol reduced delay discounting in humans, while Richards *et al*. (1999b) found no effect of alcohol on this measure; in contrast, several investigators have found impulsive choice to be induced by alcohol in rats (Tomie *et al.*, 1998; Evenden & Ryan, 1999; Hellemans *et al.*, 2005) and some also in humans (Reynolds *et al.*, 2006). Benzodiazepines have not been shown to affect impulsive choice in humans (Reynolds *et al.*, 2004a), and in different studies have been observed both to increase (Thiebot *et al.*, 1985; Cardinal *et al.*, 2000) and to decrease (Evenden & Ryan, 1996) impulsive choice in rats. These discrepancies may in some cases be due to the sensitivity of the particular task used, but may also be because the drugs (or the state of addiction) do not have a unitary effect on discounting, but one which depends heavily on the situation and the particular choices involved. For example, Mitchell has shown that cigarette deprivation increases choice impulsivity when decisions concern cigarettes, but not when they concern money (Mitchell, 2004a); likewise, smokers temporally discount cigarettes more than money (Bickel *et al.*, 1999), as well as discounting money more than controls; opiate abusers discount opiates more than money (Madden *et al.*, 1999) (Figure 6), and cocaine users discount cocaine more than money (Coffey *et al.*, 2003).

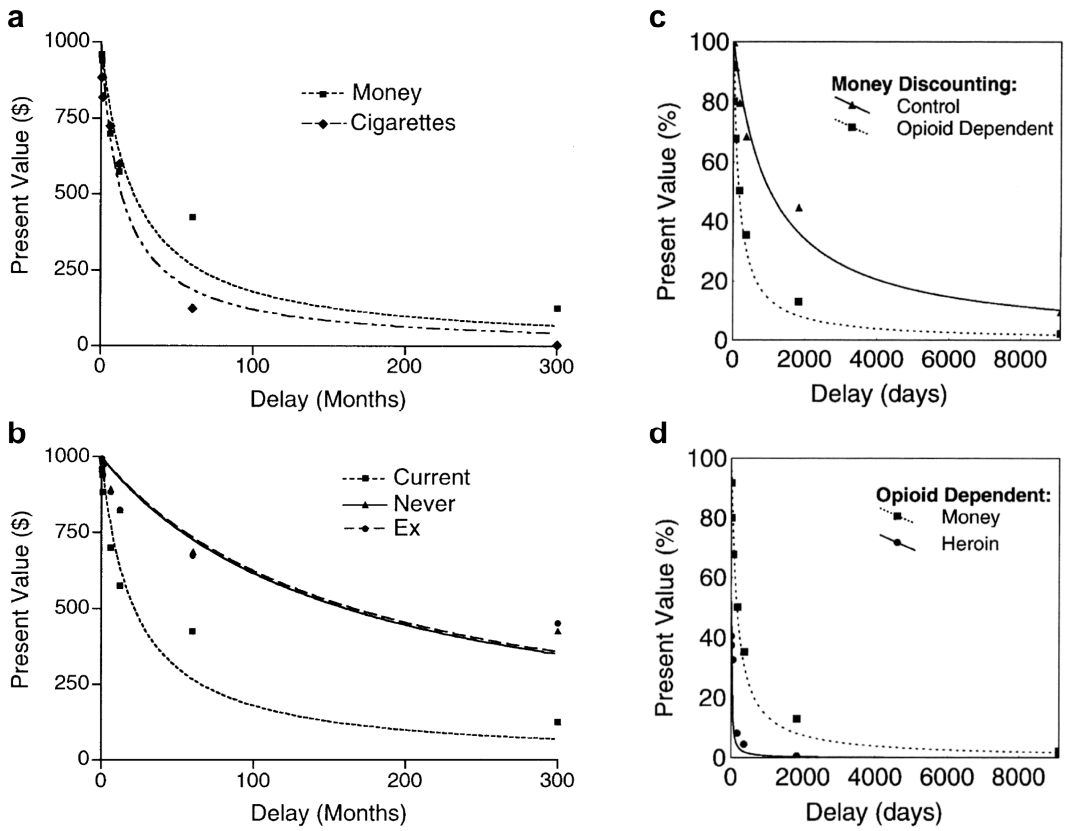

Figure 6: Exaggerated temporal discounting in drug addicts

Summary of a number of studies conducted by Bickel and colleagues examining temporal discounting in human drug addicts. **(a)** Smokers discount cigarettes slightly more steeply than money. **(b)** Current smokers discount money more than ex-smokers, or those who have never smoked, do. **(c)** Opioid-dependent humans discount money more than control subjects do. **(d)** Opioid addicts discount heroin more steeply than they discount money. Data for smokers from Bickel *et al.* (1999); data for heroin addicts from Madden *et al.* (1999), reproduced from Bickel & Johnson (2003).

1.9 ANATOMY AND CONNECTIONS OF KEY LIMBIC STRUCTURES

In this section I will outline the concept of the limbic system and in particular the anatomy of the limbic corticostriatal "loop", the striatal component of which is the nucleus accumbens. I will describe the basic anatomy and connectivity of the Acb and of the hippocampus, the two areas whose roles in delayed/uncertain reinforcement are examined in this thesis.

1.9.1 Anatomy of the limbic corticostriatal "loop"

Early investigations of the functions of hypothalamic regions (Hetherington & Ranson, 1939; Anand & Brobeck, 1951) demonstrated that electrolytic regions of the lateral hypothalamus appeared to leave animals demotivated, with impairments in unlearned behaviour (including aphagia, adipsia, and a reduction in sexual, exploratory, and maternal behaviours) and in learned behaviour (impaired instrumental responding). However, such lesions also disrupt the medial forebrain bundle, a fibre tract that passes through the lateral hypothalamus and contains the projection from midbrain DA nuclei, the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA), to the forebrain. Use of the DA-depleting toxin 6-hydroxydopamine (6-OHDA) showed that lesions of this projection or DA depletion of the striatum, one of its targets, produced a similar pattern of behavioural impairment (Stricker & Zig-

mond, 1976; Marshall & Teitelbaum, 1977); this focused attention on the role of DA and the structures that it innervated in motivated behaviour.

The basal ganglia comprise a number of subcortical nuclei, including the striatum. The striatum may be considered the "input layer" of the basal ganglia; nearly the entire neocortex projects to it (Kemp & Powell, 1971). In turn, the striatum projects to the globus pallidus, which projects via thalamic nuclei back to the cortex; the whole makes up a "loop". It is a particular characteristic of basal ganglia–thalamocortical ("corticostriatal") loops that although large areas of cortex send information into the loop, only a relatively small area of cortex is targeted by the return projection. Information flow in different loops is segregated—that is, the loops operate in parallel—and the loops are named for the areas of cortex to which they project. They are the *motor* loop (projecting in primates to the premotor cortex, supplementary motor area, and primary motor cortex, and involved in the initiation of motor acts); the *oculomotor* loop (projecting to the frontal eye fields); the *dorsolateral prefrontal* or "cognitive" loop; the *lateral orbitofrontal* loop, and the anterior cingulate or *limbic* loop (projecting to the anterior cingulate cortex and medial orbitofrontal cortex) (DeLong & Georgopoulos, 1981; Alexander *et al.*, 1986). Indeed, functional segregation (parallel processing) is apparent even within each loop (Alexander & Crutcher, 1990). The loops may also be differentiated on the basis of the parts of the basal ganglia and thalamus they pass through; thus, while inputs to the motor and cognitive loops target the dorsal striatum (caudate–putamen or neostriatum), information entering the limbic loop does so through the ventral striatum. The ventral striatum consists of the Acb, ventromedial portions of the caudate and putamen, and the olfactory tubercle; the largest component is the Acb. Within each corticostriatal loop, the basic circuitry is similar across the dorsal striatum and much of the ventral striatum (Heimer *et al.*, 1995); it is therefore likely that the various basal ganglia loops process information in qualitatively similar ways, with the nature of the cortical target determining the apparent function of each loop.

Information processing in the basal ganglia is complex, involving not only a "direct" pathway from striatum to globus pallidus (more specifically, in primates, to the internal segment of the globus pallidus and the substantia nigra pars reticulata) but a functionally antagonistic "indirect" pathway from the striatum to the globus pallidus (external segment), which projects to the subthalamic nucleus, and thence to the globus pallidus (internal segment) (Alexander & Crutcher, 1990). Cellular activity in the striatum is regulated by dopaminergic projections from the midbrain. The dorsal striatum is innervated by the SNc while the ventral striatum receives its projections from the VTA. In a further subdivision of the dorsal striatum, histochemically distinct *patches* or striosomes may be defined, which may project back to midbrain dopaminergic and cholinergic neurons, while the *matrix* circuitry is as described above (Grove *et al.*, 1986; Jiménez-Castellanos & Graybiel, 1989; Gerfen, 1992a; Gerfen, 1992b; Fallon & Loughlin, 1995), though it is not clear that this distinction applies to the ventral striatum (Heimer *et al.*, 1995). In addition, there are significant DA projections to cortical structures that provide information to, and receive information from, the basal ganglia, such as the PFC and amygdala (Fallon & Loughlin, 1995).

Here, I will focus on the limbic loop, depicted in Figure 7. Its components include many of the structures considered part of the limbic system. The term "limbic" was coined by Broca (1878) for the cortical structures encircling the upper brain stem (limbus, L. edge or border). The "limbic lobe" was suggested to have a role in emotional experience and expression by Papez (1937), concepts later to be elaborated by MacLean (1949; 1952; 1993), who introduced the expression "limbic system" to refer to the limbic lobe and its connections with the brainstem. The limbic system is not precisely defined: as the limbic lobe was considered the neural substrate for emotions, structures whose functions have to do with motivation and emotion have since been added to the anatomical definition. A modern definition of the limbic system in

primates would include cingulate and orbitofrontal cortex; the hippocampal formation, parahippocampal gyrus and mammillary bodies; anterior and medial thalamic nuclei; the nucleus accumbens and ventral pallidum; the amygdala and the hypothalamus. Key elements of the limbic corticostriatal loop are shown in Figure 7 (p. 21), with anatomically accurate depictions in Figure 8 (coronal views; p. 22), Figure 9 (sagittal views; p. 23), Figure 10 (horizontal views; p. 24), and Figure 11 ("glass brain" views; p. 25).

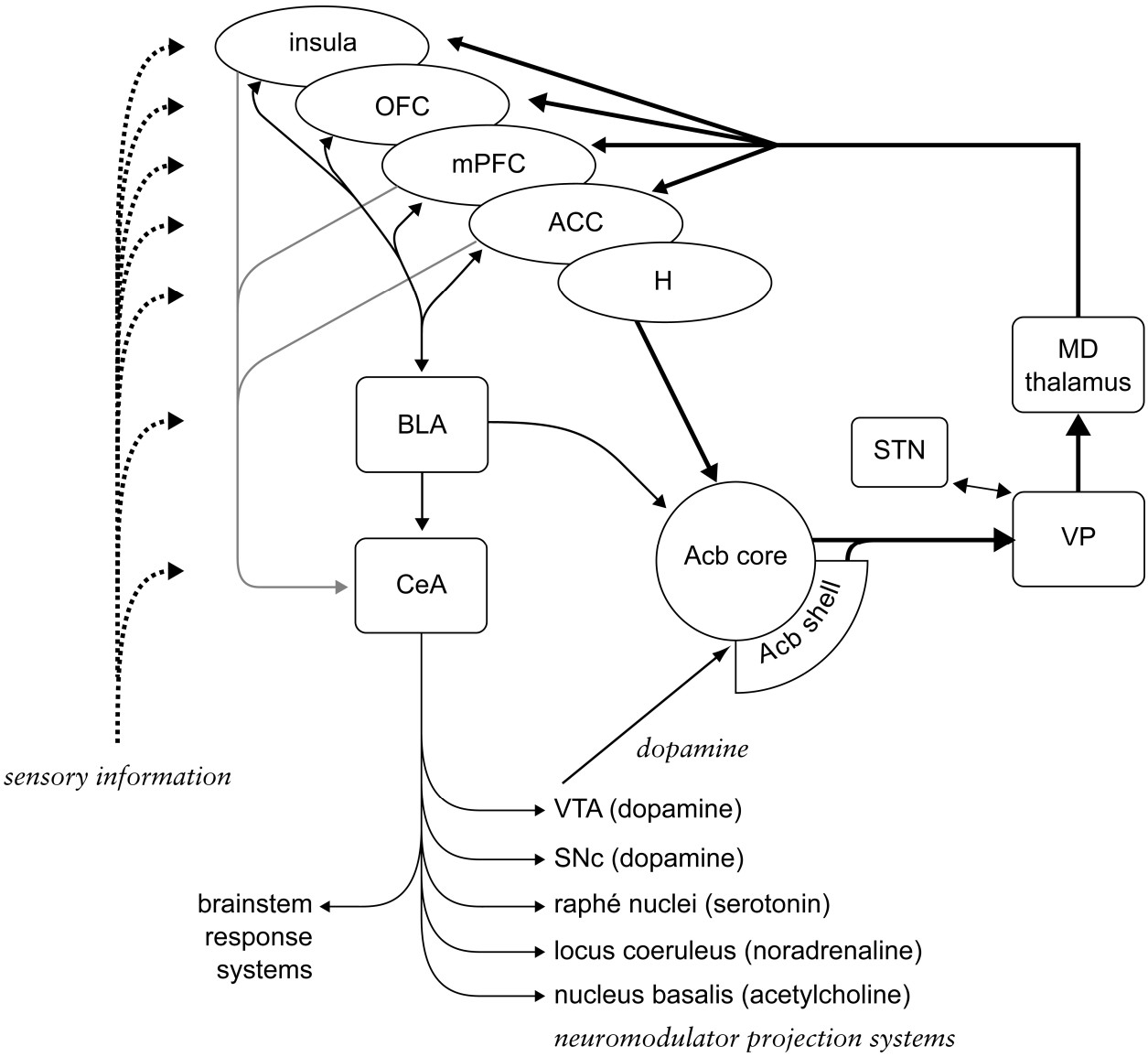

Figure 7: Key elements of the limbic corticostriatal "loop"

Simplified schematic of the limbic corticostriatal loop (after Cardinal *et al.*, 2002a), showing key structures. OFC, orbitofrontal cortex; mPFC, medial prefrontal cortex (prelimbic/infralimbic cortex in the rat); ACC, anterior cingulate cortex; H, hippocampal formation; BLA, basolateral amygdala; CeA, central nucleus of the amygdala; Acb, nucleus accumbens; STN, subthalamic nucleus; VP, ventral pallidum; MD, mediodorsal; VTA, ventral tegmental area; SNc, substantia nigra pars compacta. Not all structures and connections are shown; for example, there are projections from prefrontal cortical regions, including the OFC, to the STN (Berendse & Groenewegen, 1991; Maurice *et al.*, 1998; Hamani *et al.*, 2004).

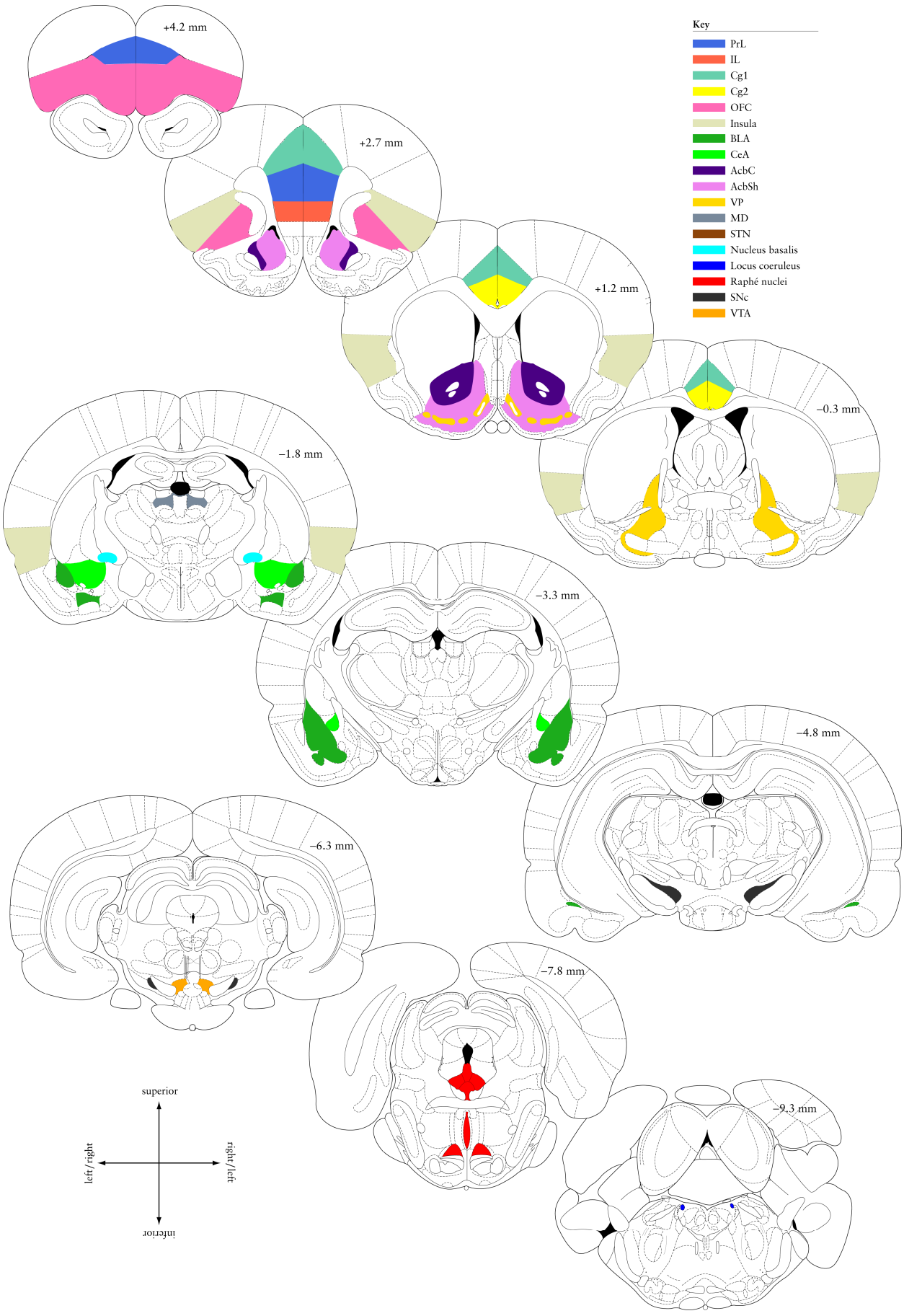

Figure 8: Coronal sections of the rat brain, showing selected limbic and related structures.
For full legend, see p. 26.

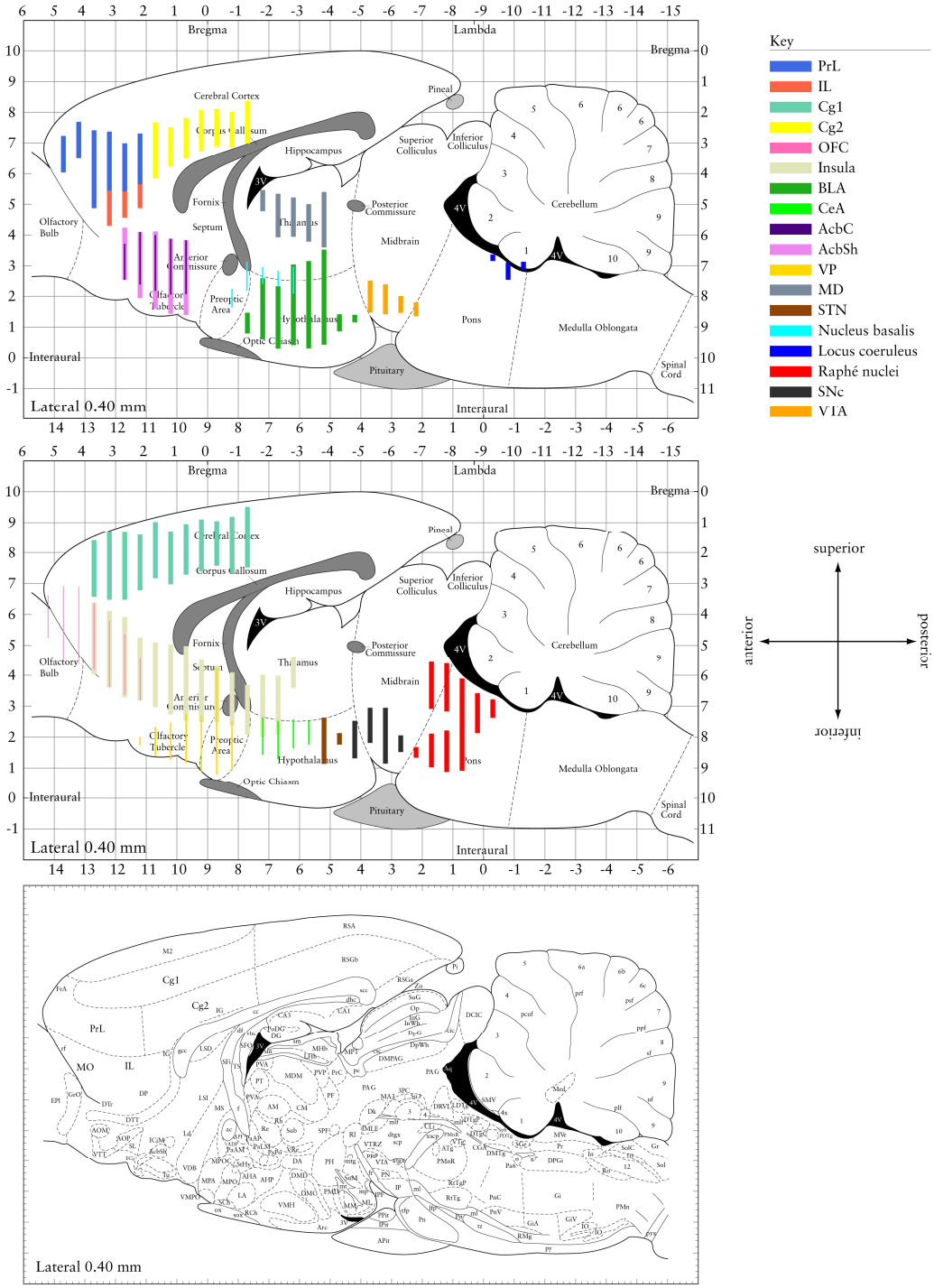

Figure 9: Sagittal paramedian views of the rat brain, showing selected limbic and related structures.
For full legend, see p. 26.

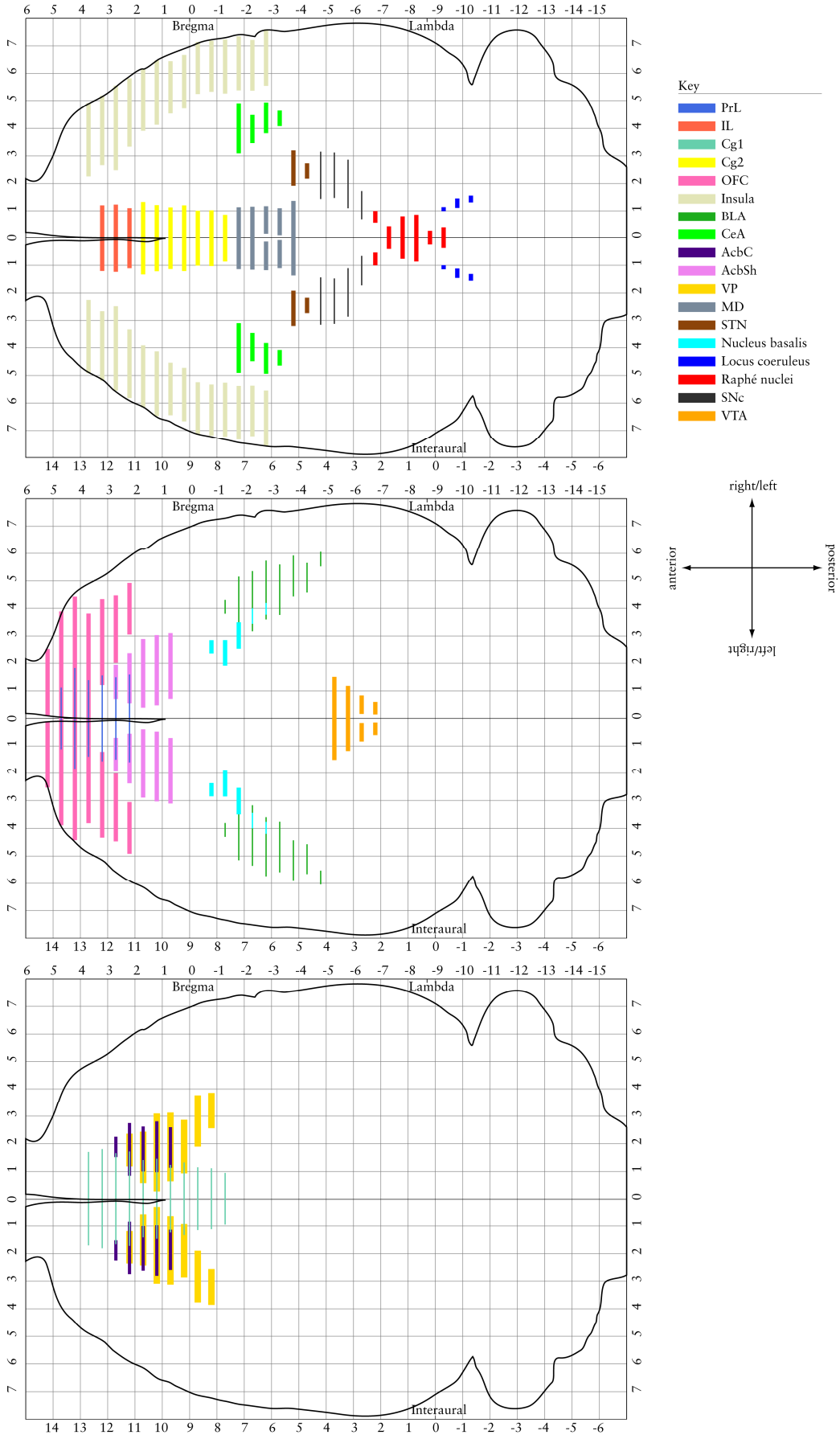

Figure 10: Horizontal views of the rat brain, showing selected limbic and related structures.
For full legend, see p. 26.

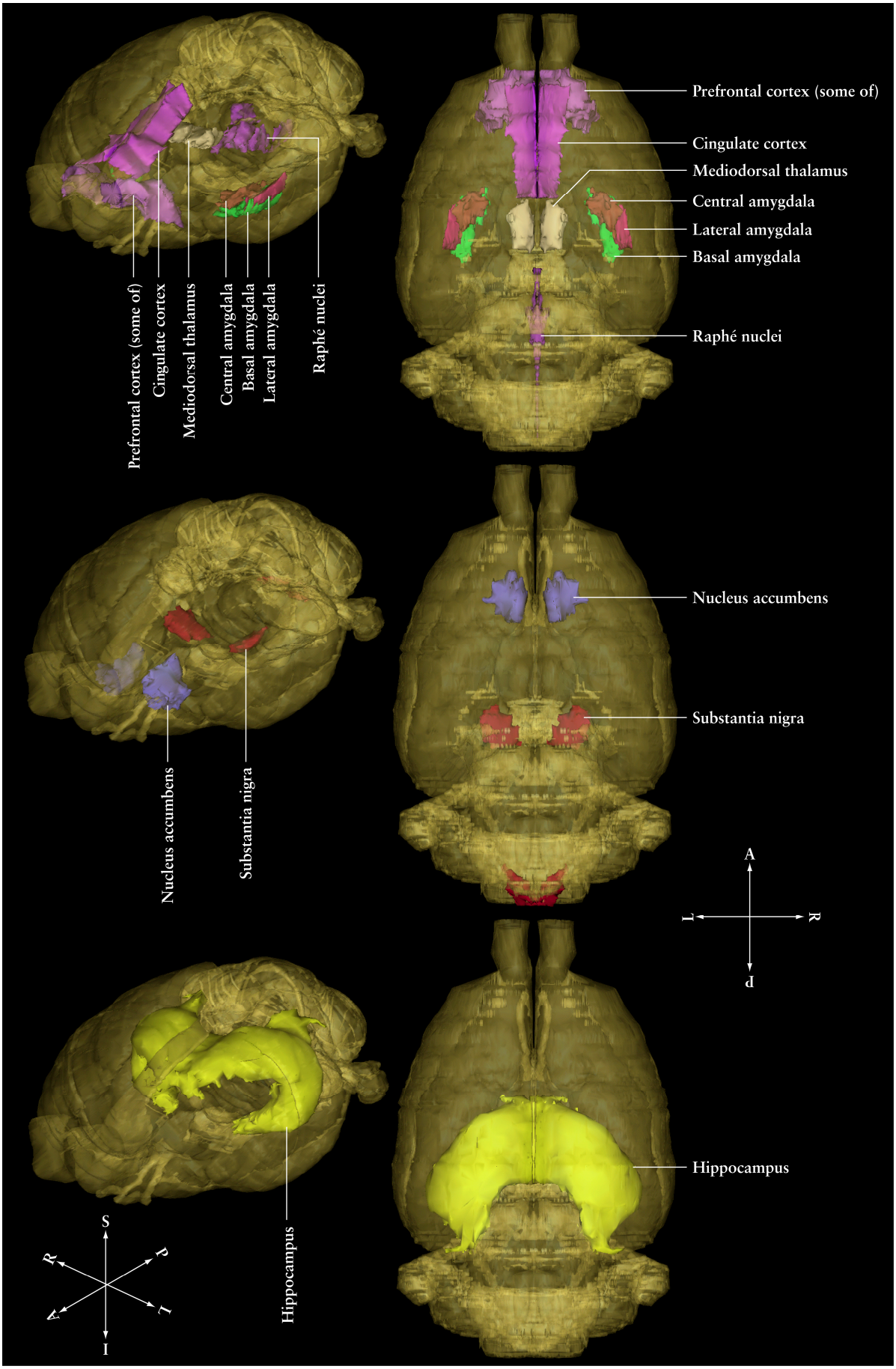

Figure 11: "Glass brain" views showing selected limbic and related structures.

For full legend, see p. 26.

Legends continued from Figure 8–Figure 11 (pp. 22–25).

Figure 8 (p. 22): Coronal sections of the rat brain, showing selected limbic and related structures.

Coronal sections are shown at 1.5 mm intervals (from +4.2 mm to –9.3 mm relative to bregma, with positive being anterior). Colours indicate selected regions of interest. Ventricles are shown in black; the SNc is shown in near-black and is located inferiorly and bilaterally at –4.8 mm and –6.3 mm slices. Coronal sections are taken from Paxinos & Watson (1997). The hippocampus is not highlighted (for which, see Figure 12). **Abbreviations and regional definitions:** PrL, prelimbic cortex; IL, infralimbic cortex; Cg1, cingulate area 1; Cg2, cingulate area 2; OFC, orbitofrontal cortex (including areas MO [medial orbital cortex], VO [ventral orbital cortex], and LO [lateral orbital cortex]). The insula, or insular cortex, includes agranular insular cortex (AI), dysgranular insular cortex (DI), and granular insular cortex (GI). BLA, basolateral amygdalar complex (including the basolateral amygdaloid nucleus BL, the basomedial amygdaloid nucleus BM, and the lateral amygdaloid nucleus La); CeA, central amygdaloid nucleus; AcbC, nucleus accumbens core; AcbSh, nucleus accumbens shell; VP, ventral pallidum; MD, mediodorsal nucleus of the thalamus; STN, subthalamic nucleus (STh in the terminology of Paxinos & Watson, 1997). "Nucleus basalis" refers to the nucleus basalis magnocellularis, or basal nucleus of Meynert (B in the terminology of Paxinos & Watson, 1997). Raphé nuclei shown include the dorsal raphé nucleus (DR), the median raphé nucleus (MnR), and the B9 group of serotonergic cells. SNc, substantia nigra pars compacta (SNC in the terminology of Paxinos & Watson, 1997); VTA, ventral tegmental area.

Figure 9 (p. 23): Sagittal paramedian views of the rat brain, showing selected limbic and related structures.

Top and middle panels: sagittal sections at 0.4 mm lateral to the midline, from Paxinos & Watson (1997), with scales in mm. Superimposed upon these sections are coloured strips at 0.5 mm intervals indicating the maximum vertical extent of each structure at that anteroposterior level, throughout the brain (not just at 0.4 mm lateral to the midline); data from coronal sections of Paxinos & Watson (1997). The groupings of structures in the upper and middle panels are not functional groupings, but are chosen to minimize overlap in the figures. **Bottom panel:** another sagittal view at 0.4 mm lateral to the midline, from Paxinos & Watson (1998), labelled. Not all abbreviations will be defined here; the picture illustrates the surface topography of prefrontal cortex (particularly areas PrL, IL, MO, Cg1, and Cg2) near the midline. Abbreviations are as in Figure 8. The hippocampus is not shown (for which, see Figure 12).

Figure 10 (p. 24): Horizontal views of the rat brain, showing selected limbic and related structures.

Horizontal sections of the rat brain, with scales in mm. The outline of the rat brain is traced from that of *MIVA* version 0.9 (Sullivan & Zhang, 2005). Superimposed upon these are coloured strips at 0.5 mm intervals indicating the maximum horizontal extent of each structure at that anteroposterior level. The groupings of structures in the three panels are not functional, but are chosen to minimize overlap in the figures. The hippocampus is not shown (for which, see Figure 12).

Figure 11 (p. 25): "Glass brain" views showing selected limbic and related structures.

Images showing views of a transparent rat brain shell ("glass brain") in two three-dimensional projections (seen from the front/left/above, and seen from directly above) with structures illustrated as defined in and rendered by *MIVA* version 0.9 (Sullivan & Zhang, 2005). S, superior; I, inferior; A, anterior; P, posterior; R, right; L, left.

1.9.2 Basic anatomy of the nucleus accumbens

The nucleus accumbens may be divided into the core (AcbC), the shell (AcbSh), and the rostral pole, a border zone with features of the other two compartments (Zaborszky *et al.*, 1985; Zahm & Brog, 1992). The pattern of innervation of these structures differs, and the Acb may be considered as having two broad functional divisions (Brog *et al.*, 1993): (1) the core, rostral pole and lateral shell; and (2) the medial shell and septal pole. Of these, the core division more closely resembles the dorsal striatum, projecting predominantly to the ventral pallidum, while the shell division also projects to subcortical structures, such as the lateral hypothalamus and periaqueductal grey, involved in the control of innate behaviours. The con-

nections of the Acb are summarized in Table 1 and Table 2. As a recipient of information from a considerable array of limbic structures that projects additionally to nuclei known to be involved in behavioural expression, the Acb has famously been suggested to represent a "limbic–motor interface" (Mogenson *et al.*, 1980).

Region in Acb	Cortical afferents	Subcortical afferents
To all/most of the nucleus accumbens	orbital cortex posterior agranular insular cortex entorhinal cortex basal amygdala hippocampal formation (via subiculum) (Note that none of these inputs is a primary or secondary sensory area or relay.)	raphé nuclei ventral tegmental area thalamic nuclei
Shell-preferential (meaning medial shell and septal pole)	dorsal peduncular cortex infralimbic cortex pyriform cortex ventral subiculum	bed nucleus of the stria terminalis hypothalamus medial amygdala lateral habenula laterodorsal tegmental nucleus sublenticular substantia innominata lateral septal nucleus locus coeruleus
Core- or rostral pole-preferential	anterior cingulate cortex medial precentral cortex dorsal and ventral prelimbic area agranular insular cortex perirhinal cortex dorsal subiculum	dorsolateral ventral pallidum subthalamic nucleus globus pallidus substantia nigra pars compacta

Table 1: Some inputs to the nucleus accumbens

Subcortical connections are nearly all reciprocal. Information from Brog *et al.* (1993) and Berendse *et al.* (1992). See Brog *et al.* (1993) for further details of thalamic connections. Table reproduced from Cardinal (2001).

Region in Acb	Efferent connections
Core	ventral pallidum subthalamic nucleus substantia nigra pars reticulata
Shell	ventral pallidum ventral tegmental area substantia nigra pars compacta hypothalamus (preoptic, medial, lateral areas) lateral septum bed nucleus of the stria terminalis lateral habenula periaqueductal grey
Indirect, via ventral pallidum	mediodorsal thalamus pedunculopontine area (part of the mesencephalic locomotor region)

Table 2: Some outputs from the nucleus accumbens

For references, see Pennartz *et al.* (1994). Table reproduced from Cardinal (2001).

1.9.3 Basic anatomy of the hippocampus

The term "hippocampus" is usually taken to mean the cornu ammonis (CA, or Ammon's horn), the dentate gyrus, and the subiculum (Aggleton & Brown, 1999). The cornu ammonis has a number of subfields, termed CA1–4. The hippocampus is archicortex. It has bidirectional links with adjacent entorhinal cortex, which itself communicates with perirhinal and parahippocampal cortex. The other main conduit of information to and from the hippocampus is via the fornix, a fibre tract that starts with its fimbriae (L. fringes)

on the hippocampus, and terminates predominantly in the mammillary bodies (part of the hypothalamus), and the anterior thalamic nuclei, but also in the nucleus accumbens. The mammillary bodies themselves project to these thalamic nuclei via the mammillothalamic tract. The macroscopic anatomy of the rat hippocampus is shown in Figure 12.

Within the hippocampus, there is a well-described trisynaptic circuit (Andersen *et al.*, 1969; 1971) (Figure 13, Figure 14). All major association areas of cortex project reciprocally to the entorhinal cortex. (1) Entorhinal cortex cells project via the *perforant path* directly to the dentate gyrus, crossing the hippocampal fissure in the process. (2) Dentate gyrus cells (specifically, granule cells) project via so-called *mossy fibres* to CA3. (3) In addition to sending axons out along the fornix, CA3 cells project via *Schaffer collaterals* to the CA1 field. After this, CA1 axons project either back to the subiculum (and from there back to entorhinal cortex) or to the fornix. The complete set of circuitry is, of course, more complex than

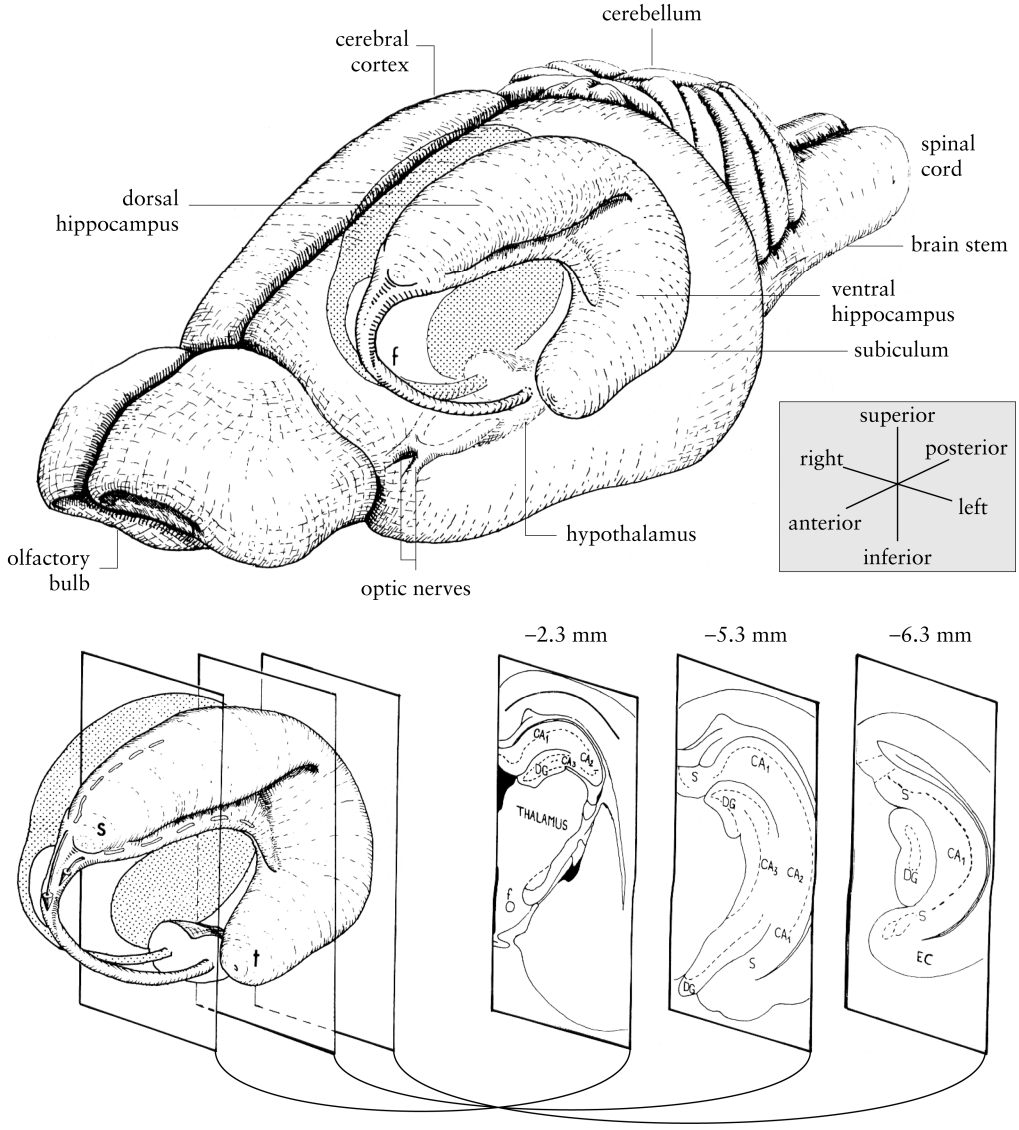

Figure 12: Diagram of the rat hippocampus

Drawings of the rat brain showing the three-dimensional organization of the hippocampus and related structures. Three coronal sections through the left hippocampus are shown at the bottom right of the figure, with their approximate anteroposterior coordinate relative to bregma. CA1, CA2, CA3: cornu ammonis fields 1–3; DG: dentate gyrus; EC: entorhinal cortex; f: fornix; s: septal pole of the hippocampus; S: subiculum; t: temporal pole of the hippocampus. Adapted from Figure 1 of Amaral & Witter (1995); copyright Elsevier 1995; reproduced in Cheung & Cardinal (2005) with permission from Elsevier.

this basic description (Figure 13). The hippocampus also receives important modulatory input, including ACh. The main forebrain cholinergic innervation comes from the nucleus basalis, which provides ACh to neocortex, and the nearby septum (septal nuclei) and diagonal band of Broca in the basal forebrain, which together provide much of the ACh input to the hippocampus. Cholinergic cells of the medial septum project via the fornix to all regions of the hippocampus; in turn, CA3 projects back to the lateral septum, where inhibitory interneurons project to the medial septum.

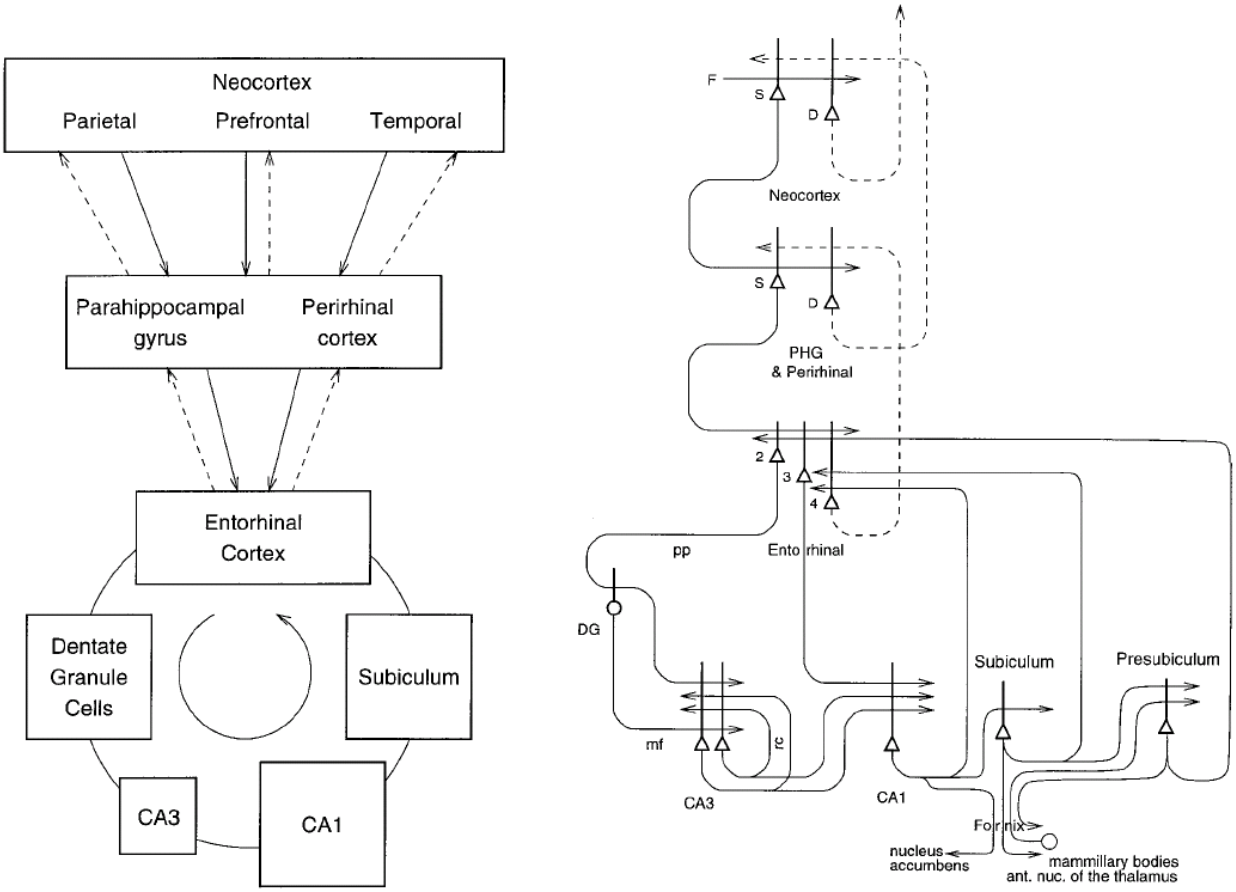

Figure 13: Outline of connections of the hippocampus

The hippocampus comprises the cornu ammonis, the dentate gyrus, and the subiculum. **Left:** basic intrinsic and extrinsic connections of the hippocampus, excluding the fornix. **Right:** synaptic connections in more detail, showing principal excitatory neurons and including fornical connections. CA, cornu ammonis; D, deep pyramidal cells; DG, dentate gyrus granule cells; F, forward inputs to association cortex areas from preceding cortical areas in the hierarchy shown at left; mf, mossy fibres; PHG, parahippocampal gyrus and perirhinal cortex; pp, perforant path; rc, recurrent collateral of the CA3 hippocampal pyramidal cells; S, superficial pyramidal cells; 2, pyramidal cells in layer 2 of the entorhinal cortex; 3, pyramidal cells in layer 3 of the entorhinal cortex. Thick lines above cell bodies represent dendrites. Figure taken from Rolls (2000).

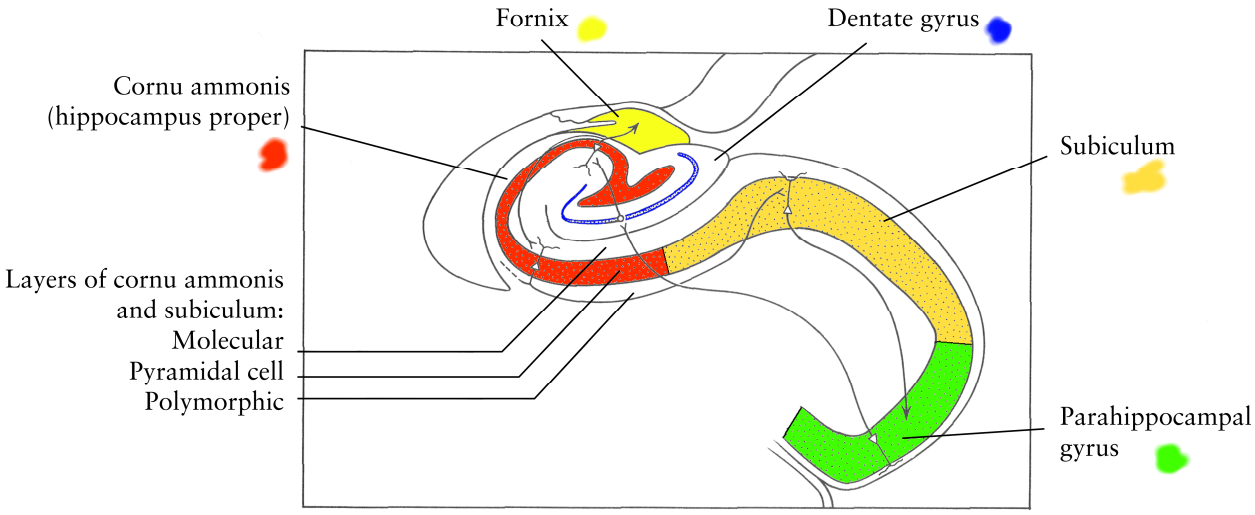

Figure 14: Cross-sectional structure of the primate hippocampus

Coronal section: medial is to the right, superior is upwards. The diagram shows the interlocking C shapes (CA fields and dentate gyrus) that typify the hippocampus, and the major pathways through the hippocampal formation. Modified from Martin (1989, p. 391). CA subfields are numbered with CA1 closest to the subiculum. In this diagram, the entorhinal cortex is considered part of the parahippocampal gyrus.

1.9.4 A note on the interpretation of excitotoxic lesion methods

Although correlative techniques such as electrophysiology and functional neuroimaging allow the functioning of the normal brain to be measured, interventional techniques (such as lesion studies or drug infusions) are required to establish a causal link between a neural structure and an aspect of behaviour. In such studies, the anatomical specificity of the method is important. The use of aspirative or radiofrequency lesions, or local anaesthetic inactivation, will destroy or inactivate neurons in the target area, but will also affect fibres (axons) passing through the target structure, potentially affecting the function of neurons whose cell bodies are elsewhere. In the present thesis, excitotoxic lesion techniques are used to affect neurons in the target site selectively. Excitotoxins typically activate *N*-methyl-D-aspartate-type (NMDA-type) glutamate receptors on neurons, leading to abnormal Ca^{2+} influx and cell death via apoptosis or excitotoxic necrosis; reviews have been provided by Choi (1988; 1995). Table 3 shows the conclu-

Manipulation	Conclusions that may be drawn from impairment	Conclusions that may be drawn from normal behaviour
Lesion, then train/test	Structure is required for learning or performance of the task.	Structure is not required for learning or performance of the task, though it may still be involved.
Train, lesion, test	Structure is required for performance of the task. Does not distinguish "mnemonic" from "motor" function.	Structure not required for performance of the task.
Train in the presence of reversible inactivation; test subsequently	Either of: (a) the structure is required for task performance, and successful performance is required as part of the learning process (e.g. instrumental behaviour); (b) the structure is involved in learning the task.	Structure not required to learn the task.
Disconnection lesion (unilateral lesion of site A and unilateral lesion of site B in the opposite hemisphere)	Site A or B must be intact bilaterally for task performance (control procedures should address this issue), or a functional connection between structures A and B is necessary for the task.	Either of: (a) a direct or indirect connection exists between the remaining A and B sites; (b) functional communication between A and B is not necessary for the task.

Table 3: Interpretation of lesion studies

Modified from Cardinal (2001).

sions that may be drawn from some of these interventional techniques.

1.10 BASIC NEUROBIOLOGY OF REINFORCEMENT LEARNING

Having outlined the psychological processes known to play a part in reinforcement learning (p. 2) and the structures of the limbic corticostriatal loop (p. 19), it will be helpful to attempt a rough correspondence before examining the specific contribution of limbic structures in the context of delayed and uncertain reinforcement. A number of limbic cortical and subcortical structures play a role in assessing the value of reinforcers and of stimuli that predict them, and in actions directed at obtaining those reinforcers or stimuli (Cardinal *et al.*, 2002a); here, I will summarize some major themes in reinforcement neuroscience.

1.10.1 Mesolimbic dopamine and the nucleus accumbens

The discovery that rats would work very hard to stimulate regions of their brain electrically—ICSS (Olds & Milner, 1954)—was historically important. Many sites that support ICSS lie on the path of dopaminergic (DAergic) neurons from the SNc and VTA to limbic sites including the ventral striatum (nucleus accumbens), and ICSS is substantially reduced after Acb DA depletion (Fibiger *et al.*, 1987). Furthermore, the rate at which rats learn an ICSS response is correlated with the degree of potentiation of synapses made by cortical afferents onto striatal neurons, a potentiation that requires DA receptors (Reynolds *et al.*, 2001). The natural idea that follows is that DA "stamps in" stimulus–response connections. Indeed, DA has acute effects to modulate corticostriatal transmission, but it also has lasting effects; most likely, the combination of cortical (presynaptic) and striatal (postsynaptic) activity normally induces long-term depression of corticostriatal synapses, but if the same pattern of activity is paired with a pulse of DA, then the active synapses are strengthened (Reynolds & Wickens, 2002). Natural reinforcers, drugs of abuse, and CSs that predict either, trigger increases in DA release in the Acb (Berridge & Robinson, 1998; Datla *et al.*, 2002; Ito *et al.*, 2002; Carelli & Wightman, 2004; Young, 2004). DA neurons fire to unexpected rewards, or unexpected stimuli that predict reward; that is, they signal reward prediction error (Schultz *et al.*, 1997; Schultz, 1998; Schultz *et al.*, 1998; Schultz & Dickinson, 2000; Schultz, 2006). DA neuron firing may be a teaching signal used for learning about actions that lead to reward (Schultz *et al.*, 1997). The Acb similarly responds to anticipated rewards (Schultz *et al.*, 1992; Miyazaki *et al.*, 1998; Martin & Ono, 2000; Schultz *et al.*, 2000; Breiter *et al.*, 2001; Knutson *et al.*, 2001; de la Fuente-Fernandez *et al.*, 2002; Cromwell & Schultz, 2003; Elliott *et al.*, 2003; McClure *et al.*, 2003a; Bjork *et al.*, 2004; Zink *et al.*, 2004). Other parameters of DA neuronal firing may signal reward uncertainty (Fiorillo *et al.*, 2003; Schultz, 2004; 2006).

An early suggestion was that Acb DA mediated the pleasurable aspects of reward (Wise, 1981; 1982; 1985), but there is good evidence against this idea. Certainly, DA is released in response to appetitive reinforcers (e.g. Fiorino *et al.*, 1993; Wilson *et al.*, 1995; Schultz *et al.*, 1997; Berridge & Robinson, 1998; Schultz, 1998; Schultz *et al.*, 1998; Schultz & Dickinson, 2000; Datla *et al.*, 2002; Ito *et al.*, 2002; Carelli & Wightman, 2004; Young, 2004), intra-Acb DA agonists are reinforcing (Phillips *et al.*, 1994), animals may titrate their drug taking to maintain high Acb DA levels (Pettit & Justice, 1989), and some aspects of naturally reinforced and drug-reinforced responding depend on Acb DA (e.g. Pettit *et al.*, 1984; Caine & Koob, 1994; Baker *et al.*, 1998; Ikemoto & Panksepp, 1999; Dickinson *et al.*, 2000; Parkinson *et al.*, 2002; Salamone & Correa, 2002; Salamone *et al.*, 2003). However, Acb DA does not mediate "pleasure" (Fibiger & Phillips, 1988; Robbins & Everitt, 1992; Berridge & Robinson, 1998; Volkow *et al.*, 1999)—though its release may correlate with activity in other systems that do—and reinforcement operates in its absence (Ettenberg *et al.*, 1982; Pettit *et al.*, 1984). Measured by microdialysis techniques, DA is also

released in response to aversive stimuli, CSs that predict them, and other salient stimuli (see e.g. Salamone, 1994; Horvitz, 2000; Young, 2004), which would be consistent with a more general motivational role. These results are not easy to reconcile with electrophysiological studies of DA neuronal firing, which have generally suggested that firing occurs in response to appetitive but not aversive stimuli. It is possible that DA neurons fire to strong, not mild, aversive events, that the DA response to aversive events is gradual rather than phasic, or that local modulation of DA release in target regions dissociates DA release from the firing of DA neurons (reviewed by Salamone, 1994; Horvitz, 2000; Joseph *et al.*, 2003).

Targets of DA neurons certainly influence instrumental learning and responding. It is not clear what structures learn from the DA teaching signal; these probably include the dorsal striatum and PFC, but much attention has focused on the Acb. Blockade of NMDA glutamate receptors in the AcbC has been shown to retard instrumental learning for food under a variable-ratio-2 (VR-2) schedule (Kelley *et al.*, 1997), as has inhibition or over-stimulation of cyclic-adenosine-monophosphate-dependent protein kinase (protein kinase A; PKA) within the Acb (Baldwin *et al.*, 2002a). Concurrent blockade of NMDA and DA D_1 receptors in the AcbC synergistically prevents learning of a VR-2 schedule (Smith-Roe & Kelley, 2000). Once the response has been learned, subsequent performance on this schedule is not impaired by NMDA receptor blockade within the AcbC (Kelley *et al.*, 1997). Furthermore, infusion of a PKA inhibitor (Baldwin *et al.*, 2002a) or a protein synthesis inhibitor (Hernandez *et al.*, 2002) into the AcbC *after* instrumental training sessions impairs subsequent performance, implying that PKA activity and protein synthesis in the AcbC contribute to the consolidation of instrumental behaviour. Thus, manipulation of the Acb can affect instrumental learning.

However, it is also clear that the Acb is not *required* for simple instrumental conditioning—but it is strongly implicated in providing "extra motivation" for behaviour, especially when such motivation is triggered by Pavlovian CSs, or when reinforcers are delayed or require substantial effort to obtain. Rats with Acb or AcbC lesions acquire lever-press responses on sequences of random ratio (RR) schedules at normal or near-normal levels (Corbit *et al.*, 2001; de Borchgrave *et al.*, 2002) and are fully sensitive to changes in the action–outcome contingency (Balleine & Killcross, 1994; Corbit *et al.*, 2001; de Borchgrave *et al.*, 2002). Thus, the Acb is not critical for goal-directed action (see Cardinal *et al.*, 2002a); rather, it appears to be critical for some aspects of motivation that promote responding for rewards in real-life situations. For example, the Acb plays a role in promoting responding for delayed rewards (Cardinal *et al.*, 2001), to be discussed later, and is required for Pavlovian CSs to provide a motivational boost to responding (Hall *et al.*, 2001; de Borchgrave *et al.*, 2002), i.e. for PIT. PIT has sometimes been termed "wanting" (Wyvell & Berridge, 2000; Wyvell & Berridge, 2001), although the term "wanting" could equally refer to the instrumental incentive value underpinning true goal-directed action. PIT can be further enhanced by injection of amphetamine into the Acb (Wyvell & Berridge, 2000) or by corticotropin-releasing hormone (CRH) acting in the AcbSh (Pecina *et al.*, 2006), and depends on DA (Dickinson *et al.*, 2000), possibly under the control of the central nucleus of the amygdala (CeA) (Hall *et al.*, 2001). CeA control of midbrain dopaminergic systems has also been demonstrated in other tasks (Lee *et al.*, 2005a; Lee *et al.*, 2005b; Holland & Gallagher, 2006). Other motivational effects of Pavlovian CSs also depend on the Acb. CSs serve as goals for behaviours (conditioned reinforcers); although lesions of Acb subregions do not prevent animals responding for conditioned reinforcement entirely (Parkinson *et al.*, 1999a), enhancement of DA neurotransmission within the Acb can boost the efficacy of conditioned reinforcement (Taylor & Robbins, 1984; 1986; Cador *et al.*, 1991; Parkinson *et al.*, 1999a). CSs that have been paired with reward also elicit approach (Brown & Jenkins, 1968); this effect also depends on the Acb (Parkinson *et al.*, 1999a; Parkinson *et al.*, 1999b; Parkinson *et al.*, 2000c) and its DA innervation

(Parkinson *et al.*, 2002). Acb DA may also be involved in learning this approach response, again perhaps under the control of the CeA (Parkinson *et al.*, 2000b; Hall *et al.*, 2001; Cardinal *et al.*, 2002b; Parkinson *et al.*, 2002; Phillips *et al.*, 2003). Acb DA also contributes directly to subjects' motivation to work hard (Ikemoto & Panksepp, 1999; Salamone & Correa, 2002; Salamone *et al.*, 2003). In naturalistic situations, rewards are frequently available only after a delay, require considerable effort to achieve, and are signalled by environmental stimuli; thus, the Acb is central to a number of processes that require motivation (Mogenson *et al.*, 1980).

This motivational process has been suggested to be particularly significant in some addictions, and modification of it may have therapeutic potential. Although DA systems are affected by drugs of abuse and natural reinforcers such as food, some abused drugs may be more potent in this regard. For example, both food and drugs of abuse increase Acb DA, but the DA response to drugs of abuse may not habituate to the same extent as that to food (Di Chiara, 1998; Di Chiara, 2002). Sensitization occurs following psychostimulant administration directly into the VTA, which induces hypersensitivity to DA in the Acb (Cador *et al.*, 1995) and enhances the response to Pavlovian CSs associated with reward (Harmer & Phillips, 1999; Taylor & Horger, 1999; Wyvell & Berridge, 2001). It is not yet clear to what extent sensitization contributes to human addiction (Sax & Strakowski, 2001), but it has been suggested that a sensitized response to drug-associated cues contributes to drug craving—that this "incentive motivational" system becomes sensitized (Robinson & Berridge, 1993). In present animal models, however, drug sensitization enhances responding for food, or responding to CSs for food (Taylor & Horger, 1999; Wyvell & Berridge, 2001; Olausson *et al.*, 2003), but in human addiction, responding for non-drug reinforcement declines relative to that for drug reinforcement (APA, 2000). Amphetamine sensitization also enhances the subsequent development of habits (Nelson & Killcross, 2006). In any case, in animal models of drug-seeking behaviour controlled by drug-associated stimuli (Everitt & Robbins, 2000), lesions of the AcbC or disruption of its glutamatergic neurotransmission reduce drug seeking (Di Ciano & Everitt, 2001; Hutcheson *et al.*, 2001b), probably by reducing the motivational impact of the CSs. DA D3 receptors are particularly concentrated in the Acb and amygdala (Sokoloff *et al.*, 1990), and D3 receptor antagonists (Vorel *et al.*, 2002; Di Ciano *et al.*, 2003) and partial agonists (Pilla *et al.*, 1999; Cervo *et al.*, 2003) reduce cue-controlled cocaine seeking or relapse to cocaine taking in animal models. Some manipulations that reduce drug seeking or reinstatement of drug taking in animal models, such as DA D3 receptor antagonists, do not reduce food seeking in a similar manner (Vorel *et al.*, 2002; Di Ciano *et al.*, 2003).

1.10.2 Habits and the dorsal striatum

The development of motor habits may depend on dorsal striatal plasticity (Packard & McGaugh, 1996), which may in turn depend on DA receptors (Reynolds *et al.*, 2001; Reynolds & Wickens, 2002). Expression of S–R habits requires the dorsal striatum (Packard & McGaugh, 1996; Yin *et al.*, 2004), and the balance between habits and goal-directed behaviour may also be regulated by the prelimbic and infralimbic cortex (Killcross & Coutureau, 2003), subdivisions of the rat PFC. Dorsal striatal DA release is also a correlate of well-established cocaine seeking (Ito *et al.*, 2002).

1.10.3 Action–outcome contingency knowledge, planning and value: the PFC and amygdala

The PFC (specifically, prelimbic cortex) is required for rats to represent the contingencies between actions and their outcomes (Balleine & Dickinson, 1998; Corbit & Balleine, 2003), and acquisition of instrumental responses on a simple schedule is also disrupted synergistically by concurrent blockade of

NMDA and DA D_1 receptors in the PFC (Baldwin *et al.*, 2002b). The PFC is also involved in extinction (Myers & Davis, 2002)—the cessation of responding when a CS or response is no longer paired with reinforcement. Extinction is not "unlearning" but involves the learning of new, inhibitory ("CS → not-US" or "CS → don't respond") associations (see Mackintosh, 1974; Delamater, 2004). Lesions of the ventral medial PFC interfere with the extinction of Pavlovian conditioned freezing in the rat (Morgan *et al.*, 1993; Morgan & LeDoux, 1995; Morgan & LeDoux, 1999). The PFC interacts with the amygdala, an important site of CS–US association in this task (see Davis, 2000; LeDoux, 2000), and may suppress conditioned freezing when it is no longer appropriate (see Garcia *et al.*, 1999; Myers & Davis, 2002; Quirk *et al.*, 2003; Rosenkranz *et al.*, 2003).

The orbitofrontal cortex (OFC) is part of the PFC with a particular role in the assessment of reinforcer value; it has bidirectional connections to the amygdala, and both are heavily implicated in the retrieval of the value of primary reinforcers based on information from CSs (see Cardinal *et al.*, 2002a; Balleine *et al.*, 2003; Lindgren *et al.*, 2003; Pickens *et al.*, 2003; O'Doherty, 2004). The OFC may encode the economic value of goods directly (e.g. Padoa-Schioppa & Assad, 2006). In humans, the OFC and amygdala are also activated during extinction of Pavlovian conditioning (Gottfried & Dolan, 2004). The amygdala regulates the DA signal to the Acb (Everitt *et al.*, 2000; Parkinson *et al.*, 2000a; Hall *et al.*, 2001; Cardinal *et al.*, 2002a; Phillips *et al.*, 2003). Goal-directed action requires that action–outcome contingencies interact with the incentive value of goals (Dickinson, 1994; Dickinson & Balleine, 1994); the connection between the amygdala and the PFC (Pitkänen, 2000) may provide this functional link (Coutureau *et al.*, 2000; Arana *et al.*, 2003; Gottfried *et al.*, 2003; Holland & Gallagher, 2004). The retrieval of incentive value about food reinforcers also requires the gustatory cortex, the insula (Balleine & Dickinson, 1998; Balleine & Dickinson, 2000).

1.10.4 Hedonic assessment

Hedonic assessment of rewards themselves ("liking" or "pleasure"), does not depend on dopaminergic processes (Pecina *et al.*, 1997; Berridge & Robinson, 1998; Dickinson *et al.*, 2000; Pecina *et al.*, 2003). Instead, it involves opioid mechanisms in the AcbSh and other systems in the pallidum and brainstem (Berridge, 2000; Kelley & Berridge, 2002). Intra-Acb μ opioid agonists also affect food preference, increasing the intake of highly palatable foodstuffs including fat, sweet foods, salt, and ethanol (Zhang *et al.*, 1998; Zhang & Kelley, 2000; Kelley *et al.*, 2002; Zhang & Kelley, 2002; Will *et al.*, 2003; Ward *et al.*, 2006). The effect can be two-way, with chronic ingestion of chocolate inducing adaptations in endogenous Acb opioid systems (Kelley *et al.*, 2003).

1.10.5 The hippocampus and the representation of context

Interest in the hippocampus as a memory store stemmed from early observations of human amnesia following medial temporal lobe resection (Scoville & Milner, 1957; Corkin *et al.*, 1997) and many subsequent animal models (for recent reviews, see Squire, 1992; Murray & Mishkin, 1998; Baxter & Murray, 2001a; Baxter & Murray, 2001b; Clark *et al.*, 2001b), together with the discovery of synaptic long-term potentiation (LTP), a cellular mechanism of memory, in the hippocampus (Bliss & Lømo, 1973; Morris, 1994). Human anterograde amnesia has also resulted from damage to diencephalic structures, and it has been suggested that a circuit involving the hippocampus, mammillary bodies, and anterior thalamic nuclei is essential for episodic memory formation (Delay & Brion, 1969; Aggleton & Brown, 1999). There is substantial contemporary debate on the exact type of memory supported by the hippocampus and adjacent cortical regions, and the manner in which they do so (Gaffan & Harrison, 1989; Gaffan, 1992; Morris &

Frey, 1997; Eichenbaum *et al.*, 1999; Griffiths *et al.*, 1999; Morris, 2001; Good, 2002; Day *et al.*, 2003; Fortin *et al.*, 2004). However, there is good evidence that the hippocampus contributes to the representation of context.

The idea that the hippocampus plays a role in contextual representations developed from the original discovery of cells in the rat hippocampus that increased their firing rate when the rat was at a particular location in its environment—"place cells" (O'Keefe & Dostrovsky, 1971). O'Keefe & Nadel (1978) suggested that the hippocampus functions as a "cognitive map", informing the rat where it is in the world (recently reviewed by Eichenbaum *et al.*, 1999). Lesion studies support the idea that the hippocampus is critical in navigation. For example, Morris *et al.* (1982) showed that rats with hippocampal lesions were impaired at a task in which they had to learn the location of a hidden submerged platform in a tank full of opaque liquid, now known as the Morris water maze. The deficit appears to depend on navigating relative to a constellation of cues in the room, as hippocampal lesions do not impair the ability to head in a particular direction to a stimulus that bears a fixed relation to the platform (Pearce *et al.*, 1998). Water maze performance is damaged by dorsal, not ventral hippocampal lesions (Moser *et al.*, 1995). Learning in the water maze can be blocked by the glutamate NMDA receptor antagonist D-(–)-2-amino-5-phosphonopentanoic acid (AP-5), which blocks LTP (Morris *et al.*, 1986); similar effects follow NMDA receptor subunit mutations. However, the effects of AP-5 are attenuated if the rats are trained in a different water maze beforehand (Bannerman *et al.*, 1995), so the role of the NMDA receptors may not be a specifically spatial one. Human imaging studies also support the idea of a role in the hippocampus in navigation (e.g. Maguire *et al.*, 1997; Maguire *et al.*, 1998; Maguire *et al.*, 2000).

It has been argued that the hippocampus doesn't encode a map in the conventional sense; rather, it appears that place cells encode the relationship between subsets of cues in the environment, independent of other cues (Eichenbaum *et al.*, 1999). Hippocampal neurons also encode nonspatial features (Wood *et al.*, 1999). Eichenbaum *et al.* (1999) suggest that the hippocampus can encode spatial information because this is a special case of encoding the relations between stimuli. These relations are useful for navigation when they are spatial relations, but the memories encoded by the hippocampus can be used for other purposes. The use of a more abstract relationship is demonstrated by transitive inference. If a subject learns that B > C (where ">" denotes "should be chosen over") and C > D, then the logical property of transitivity should allow it to infer that B > D. Dusek & Eichenbaum (1997) have shown that fornix transection and perirhinal/entorhinal cortex lesions, both of which partially disconnect the hippocampus, impair transitive inference in rats. Similarly, Eichenbaum and colleagues have demonstrated that the hippocampus contributes to the memory for sequences of events in rats (Fortin *et al.*, 2002; Ergorul & Eichenbaum, 2006).

Both the hypothesis that the hippocampus encodes spatial relationships (e.g. Eichenbaum *et al.*, 1999), and the hypothesis that it encodes visual scenes (Gaffan, 1992), predict that the hippocampus might be involved in associating together the many visual and non-visual elements that make up a spatial environment, or context. Therefore, it might be expected that the hippocampus contributes to contextual conditioning. In a prototypical task, if a rat receives tone–shock pairings in a distinct environment, it may subsequently show "fearful" reactions to the tone (discrete CS conditioning) and also the environment (contextual conditioning). Lesions of the hippocampus have been shown to impair Pavlovian conditioning to a contextual CS, but not to a discrete CS, in rats (Hirsh, 1974; Selden *et al.*, 1991; Kim & Fanselow, 1992; Phillips & LeDoux, 1992; Honey & Good, 1993; Jarrard, 1993; Kim *et al.*, 1993; Phillips & LeDoux, 1994; Phillips & LeDoux, 1995; Chen *et al.*, 1996; Maren & Fanselow, 1997; Anagnostaras *et al.*, 1999; Rudy *et al.*, 2002), at least for some processes involving contextual representation (Good & Honey, 1991;

Holland & Bouton, 1999; Good, 2002). Context-specific neuronal firing patterns develop in the hippocampus of rats required to discriminate different contexts (Smith & Mizumori, 2006). In some cases, discrete CS conditioning has even been enhanced following hippocampal lesions (e.g. Ito *et al.*, 2005), which may reflect a reduction in contextual competition (see p. 8).

1.11 NEUROANATOMICALLY SPECIFIC STUDIES OF DELAYED OR UNCERTAIN REINFORCEMENT

In recent years, a number of studies have examined the effects of focal excitotoxic or neurochemical lesions on choice and learning involving delayed or uncertain rewards, in additional to correlational studies using functional imaging, microdialysis, and electrophysiology. These studies centre on interconnected structures forming part of the limbic corticostriatal loop (Figure 7). Initial work focused on the Acb and two of its cortical afferents, the anterior cingulate cortex (ACC) and the medial prefrontal cortex (mPFC), as structures potentially involved in regulating choice between alternative reinforcers, for three main reasons.

First, these structures have been firmly implicated in reinforcement processes. The Acb, once suggested to mediate the reinforcing efficacy of natural and artificial rewards (see Koob, 1992) (and also Wise, 1981; 1982; 1985; 1994)**,** is now thought not to be necessary for this, but instead to be a key site for the motivational impact of impending rewards (reviewed by Robbins & Everitt, 1996; Salamone *et al.*, 1997; Everitt *et al.*, 1999; Parkinson *et al.*, 2000a; Cardinal *et al.*, 2002a). Many of its afferents have also been shown to be involved in reward-related learning, including the ACC (Bussey *et al.*, 1997a; Bussey *et al.*, 1997b; Parkinson *et al.*, 2000c; Cardinal *et al.*, 2003a) and the mPFC (e.g. Balleine & Dickinson, 1998; Richardson & Gratton, 1998; Bechara *et al.*, 1999; Tzschentke, 2000).

Second, these regions are important recipients of dopaminergic and serotonergic afferents (Fallon & Loughlin, 1995; Halliday *et al.*, 1995), and pharmacological manipulations of DA and 5-HT systems have been shown to affect impulsive choice in rats, as described above.

Third, abnormalities of these regions have been detected in humans with ADHD, and in animal models of ADHD. Abnormal functioning of prefrontal cortical regions, including medial prefrontal and anterior cingulate cortex, has been observed in ADHD patients (Ernst *et al.*, 1998; Bush *et al.*, 1999; Rubia *et al.*, 1999). In the SHR, differences in DA receptor density and gene expression have been observed within the core and shell regions of the Acb (Papa *et al.*, 1996; Carey *et al.*, 1998; Papa *et al.*, 1998; Sadile, 2000). Abnormalities of DA release have been detected in the Acb (de Villiers *et al.*, 1995; Russell *et al.*, 1998; Russell, 2000) and PFC (Russell *et al.*, 1995), in addition to possible dysfunction in the dorsal striatum and amygdala (Russell *et al.*, 1995; Papa *et al.*, 2000).

These early studies, described below, indicated a role for the AcbC in choosing delayed rewards; subsequent work has attempted to delineate the contribution of structures connected to it, and these will be reviewed in turn.

1.11.1 Nucleus accumbens core (AcbC)

1.11.1.1 Choice involving delayed reinforcement

The Acb responds to anticipated rewards in a variety of species (Schultz *et al.*, 1992; Miyazaki *et al.*, 1998; Martin & Ono, 2000; Schultz *et al.*, 2000; Breiter *et al.*, 2001; Knutson *et al.*, 2001; Cromwell & Schultz, 2003; Bjork *et al.*, 2004; Izawa *et al.*, 2005). As discussed above, it is innervated by DA neurons that respond to errors in reward prediction in a manner appropriate for a teaching signal (Schultz *et al.*,

1997; Schultz, 1998; Schultz *et al.*, 1998; Schultz & Dickinson, 2000; Schultz, 2006), interventional studies have shown it to be a key site for the motivational impact of impending rewards (reviewed by Robbins & Everitt, 1996; Salamone *et al.*, 1997; Everitt *et al.*, 1999; Parkinson *et al.*, 2000a; Cardinal *et al.*, 2002a; Robbins *et al.*, 2005), and Acb abnormalities have been observed in rat models of ADHD (de Villiers *et al.*, 1995; Papa *et al.*, 1996; Carey *et al.*, 1998; Papa *et al.*, 1998; Russell *et al.*, 1998; Russell, 2000; Sadile, 2000).

Causal experimental studies have shown that excitotoxic lesions of the AcbC produce impulsive choice, reducing rats' preference for large/delayed rewards, compared to small/immediate rewards (Cardinal *et al.*, 2001; 2003b). These studies used a task in which rats were offered regular choices between a one-pellet immediate reward and a four-pellet reward delayed from 0–60 s (Figure 15). No cues were present during the delay, to avoid any potential confounds arising from conditioned reinforcement effects (Cardinal *et al.*, 2000), and subjects were trained preoperatively, assigned to matched groups, operated upon, and retested postoperatively, to avoid any possible effects of the lesion on learning of the task. AcbC-lesioned subjects (Figure 16) were rendered impulsive in their choices: they exhibited a profound deficit in their ability to choose a delayed reward, and persisted in choosing impulsively even though they were made to experience the larger, delayed alternative at regular intervals (Figure 17). This effect was not due to an inflexible bias away from the lever producing the delayed reinforcer: AcbC-lesioned rats still chose the large reinforcer more frequently at zero delay than at other delays, and removal of the delays resulted in a rapid and significant increase in the rats' preference for the large reinforcer. Thus, the pattern of choice reflected a reduced preference for the large reinforcer when it was delayed, suggesting that delays reduced the effectiveness or value of reinforcers much more in AcbC-lesioned rats than in controls.

Although a few lesioned subjects avoided the large-reinforcer alternative postoperatively even when the delay was zero, this was probably due to within-session generalization from trial blocks at which delays were present (Evenden & Ryan, 1996; Cardinal *et al.*, 2000), as prolonged training in the absence of delays restored a near-absolute preference for the large reinforcer in the majority of subjects—who were then much more impulsive than shams again when delays were re-introduced (Figure 18) (Cardinal *et al.*, 2003b). These results indicate that AcbC-lesioned rats were able to discriminate the two reinforcers, but preferred immediate small rewards to larger delayed rewards.

Similar effects on preference are observed following lesions of the caudal lobus parolfactorius in the chick, thought to be the avian counterpart of the Acb (Izawa *et al.*, 2003).

Recently, AcbC lesions have also been found to impair performance on a task requiring rats to choose between an uncertain immediate reward and a certain delayed reward (Pothuizen *et al.*, 2005). One alternative required completion of a fixed-ratio-5 (FR-5) response for immediate delivery of a food pellet with probability $P = 0.25$; the other required completion of an FR-5 response for delayed certain delivery of an identical food pellet. AcbC lesions reduced rats' preference for the delayed, certain alternative, following sustained testing (Pothuizen *et al.*, 2005). AcbC lesions also reduced efficiency (the number of responses made per reward earned) in a differential-reinforcement-of-low-rates (DRL) schedule (Pothuizen *et al.*, 2005), in which animals must respond below a certain rate in order to obtain reward. This is much like the effects of whole-Acb lesions (Reading & Dunnett, 1995), although the DRL task may also be susceptible to changes in general levels of motor activity: AcbC-lesioned rats are hyperactive (Maldonado-Irizarry & Kelley, 1995; Parkinson *et al.*, 1999a; Cardinal *et al.*, 2001), and hyperactivity would itself tend to reduce DRL efficiency.

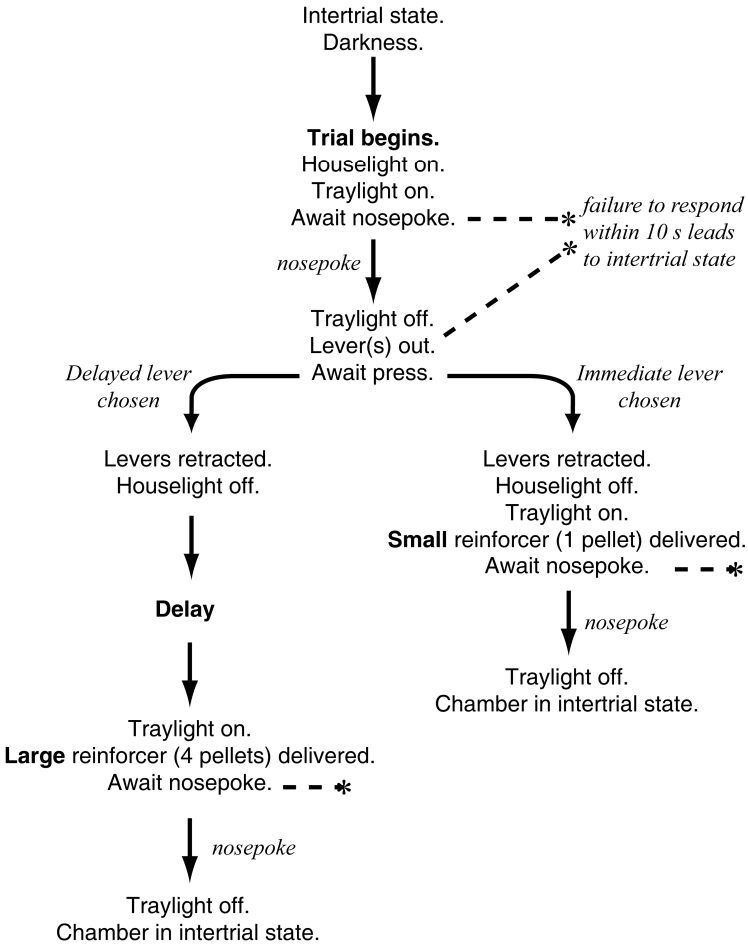

Figure 15: Task schematic: choice between small, immediate and large, delayed rewards

Hungry rats regularly choose between two levers. Responding on one lever leads to the immediate delivery of a small food reward (1 pellet); responding on the other leads to a much larger food reward (4 pellets), but this reward is delayed for between 0 and 60 seconds. The figure shows the format of a single trial; trials begin at regular intervals (every 100 s), so choice of the small reinforcer is always suboptimal. Sessions consist of 5 blocks. In each block, two single-lever trials are given (one trial for each lever), to ensure the animals sample the options available at that time; these are followed by ten choice trials. The delay to the large reinforcer is varied systematically across the session: delays for each block are 0, 10, 20, 40, and 60 s respectively. In the so-called "signalled" or "cue" condition, a stimulus light is illuminated during the delay to the large reinforcer; this is absent in the "unsignalled" or "no cue" condition, used for lesion studies. From Cardinal *et al.* (2000; 2001); based on a task by Evenden & Ryan (1996).

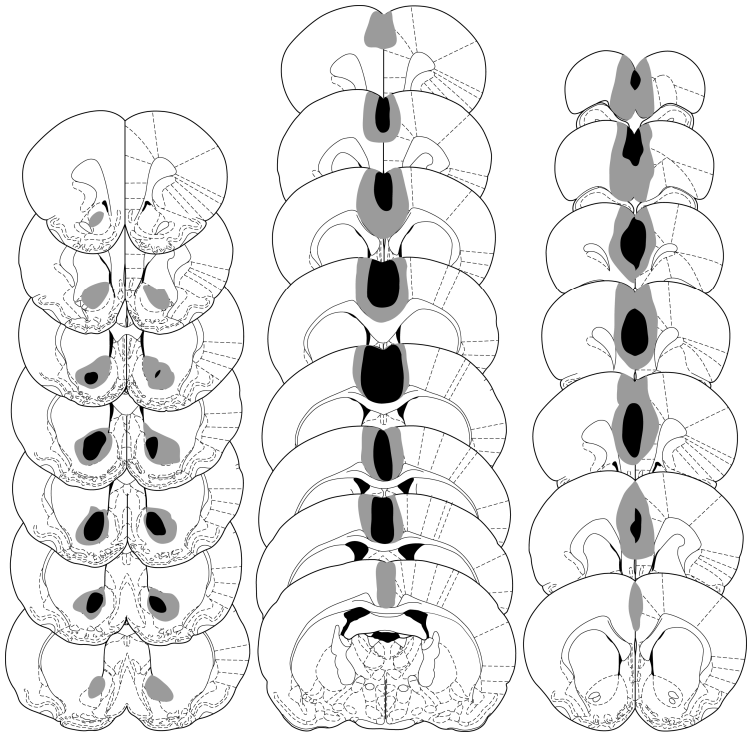

Figure 16: Schematics of lesions of the nucleus accumbens core (AcbC), anterior cingulate cortex (ACC), and medial prefrontal cortex (mPFC)

Schematics of lesions of the AcbC **(left)**, ACC **(middle)**, and mPFC **(right)**. Black shading indicates the extent of neuronal loss common to all subjects; grey indicates the area lesioned in at least one subject. Coronal sections are +2.7 through +0.48 mm (AcbC), +2.7 mm through –1.3 mm (ACC), and +4.7 through +1.7 mm (mPFC) relative to bregma. Outlines are taken from Paxinos and Watson (1998). Figure from Cardinal *et al.* (2001).

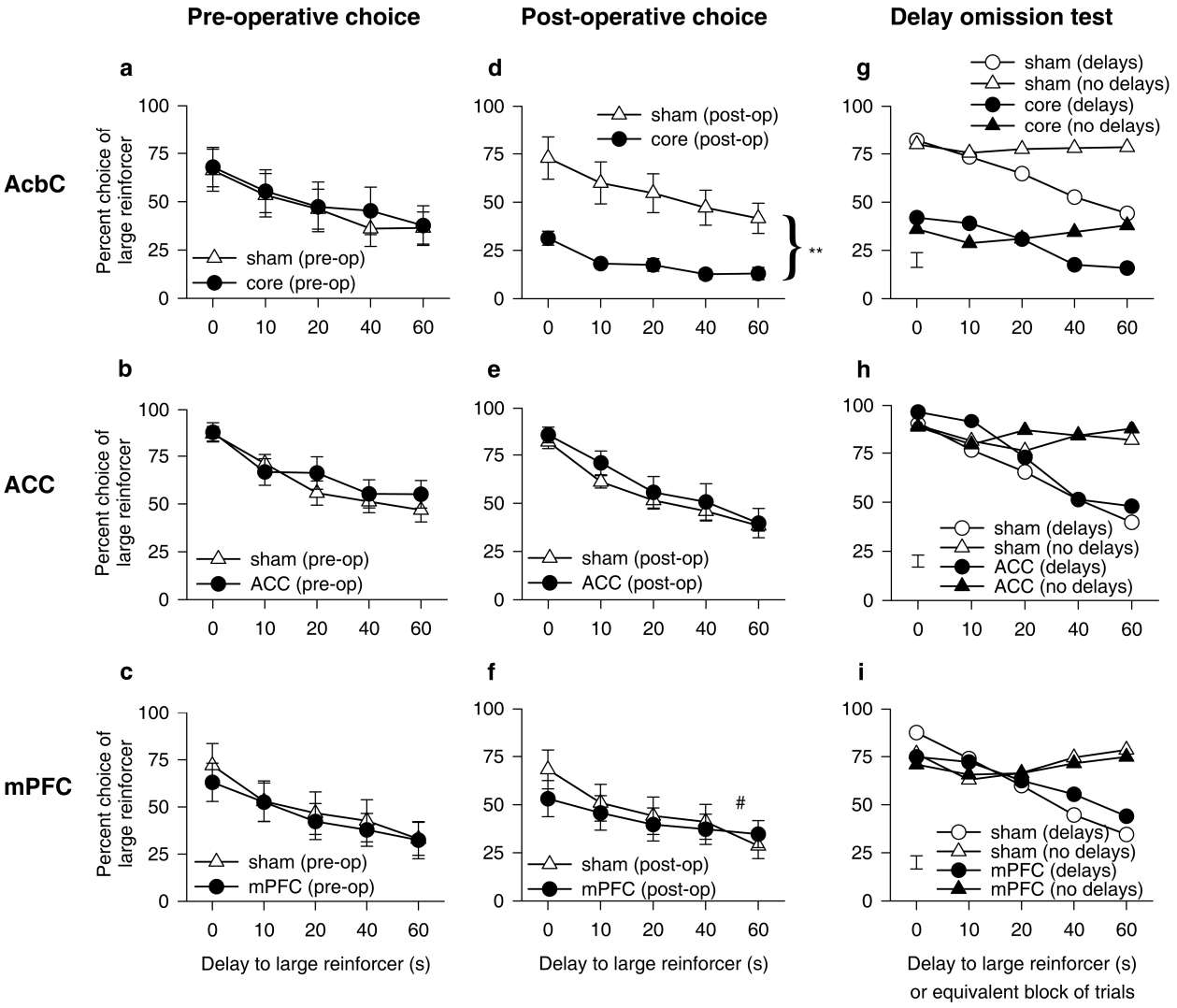

Figure 17: Choice between immediate, small and large, delayed rewards in rats with lesions of the AcbC, ACC, or mPFC

Effect of lesions of the AcbC (top), ACC (middle), or mPFC (bottom) on choice of delayed reward (● lesioned group; Δ corresponding sham group; error bars, SEM). The "no cue" condition (see Figure 15) was used throughout. **Panels a–c** show the pattern of choice in the last 3 sessions preceding surgery; corresponding sham/lesion groups were matched for performance. Subjects' preference for the large reinforcer declined with delay, as is typical for trained subjects performing this task (Evenden & Ryan, 1996; Cardinal *et al.*, 2000). **Panels d–f** illustrate choice in the first 7 postoperative sessions. The AcbC-lesioned group was markedly impaired (** $p < .01$), choosing the delayed reinforcer significantly less often than shams at every delay, including zero. However, both groups still exhibited a within-session shift in preference. ACC lesions had no effect on choice. The mPFC-lesioned subjects exhibited a "flatter" within-session preference shift than shams (# $p < .05$, group × delay interaction). **Panels g–i** illustrate the effects of omitting all delays in alternating sessions (●/○, lesioned/sham groups with delays; ▲/Δ, lesioned/sham groups without delays; error bars, SED for the three-way interaction). All groups remained sensitive to the contingencies. Delay removal increased both the sham- and AcbC-lesioned groups' preference for the larger reward; ACC- and mPFC-lesioned rats were also as sensitive to removal of the delays as shams. From Cardinal *et al.* (2001).

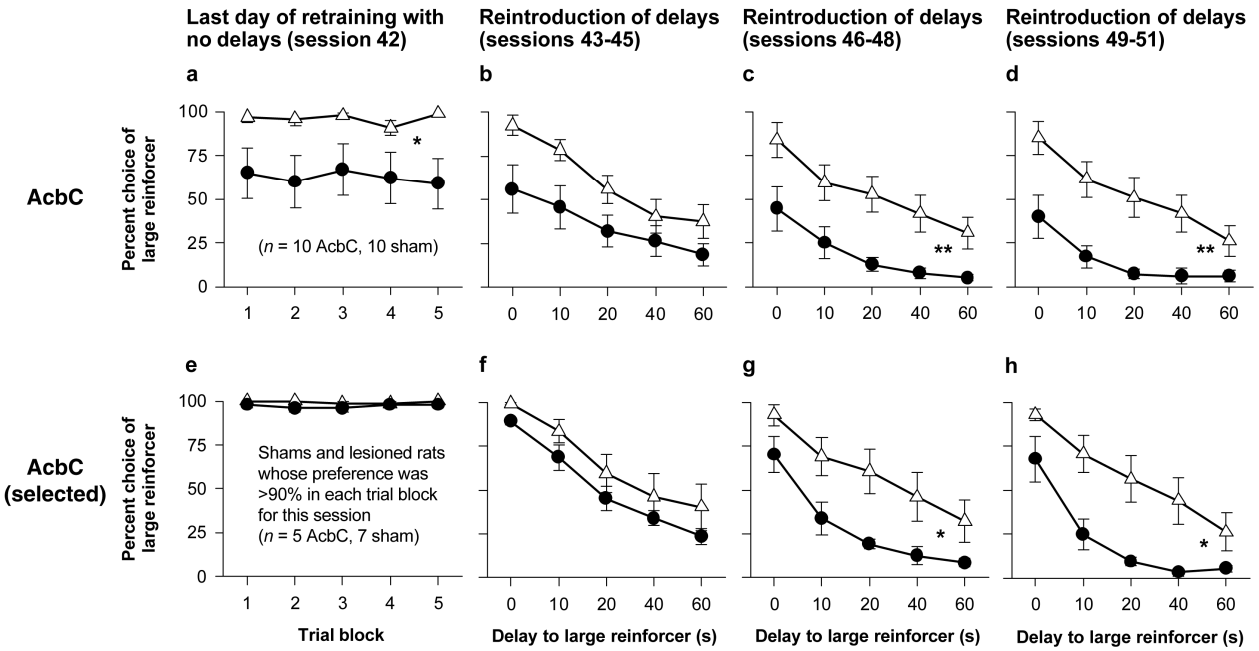

Figure 18: Further testing of AcbC-lesioned rats in the delayed reinforcement choice task

Panel a illustrates the preference of AcbC-lesioned rats following extended training in the absence of any delays (a further six sessions after completion of other behavioural tests; ● AcbC-lesioned group; Δ shams; error bars, SEM). **Panels b–d** show performance over consecutive blocks of sessions upon the reintroduction of delays (* $p < .05$, ** $p < .01$, difference from shams). **Panels e–h** show data from the same sessions as A–D, but only include data from those rats selected for ≥90% preference for the large reinforcer in every trial block on the last day of training with no delays. The sham and lesioned groups were therefore matched in E. Panels F–H show that despite this matching, preference for the large reinforcer in the AcbC group collapsed upon reintroduction of the delays. As these data exhibit significant heterogeneity of variance, the highly conservative correction of Box (1954) was applied (see Howell, 1997, pp. 322/457/464); * $p < .05$ for the corrected between-group difference. The subjects were the same as those reported in Cardinal *et al.* (2001); data from Cardinal *et al.* (2003b).

1.11.1.2 Processing of reward magnitude

Is the impulsive choice seen in AcbC-lesioned rats (Cardinal *et al.*, 2001) due to an effect on subjects' processing of reward delay, or of reward magnitude? As this task involves choice between reinforcers that differed in both magnitude and delay, impulsive choice might arise as a result of altered sensitivity to reinforcer magnitude, or delay, or both (Ho *et al.*, 1999) (Figure 19). Lesioned rats might have chosen the immediate small reward because they did not perceive the large reward to be as large (relative to the small reward) as sham-operated controls did, in which case the abnormally low magnitude of the large reward would be insufficient to offset the normal effects of the delay. Alternatively, they might have perceived the reward magnitudes normally, but been hypersensitive to the delay.

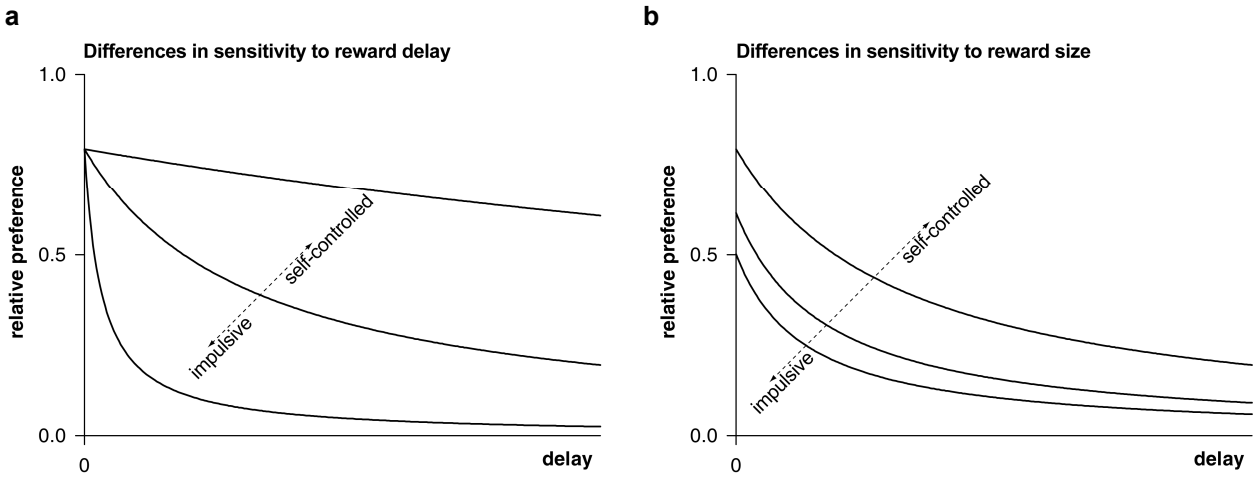

Figure 19: Delay and magnitude discounting applied to choice

Differences in delay discounting and differences in magnitude discounting can both affect choice between a smaller/sooner (SS) and a larger/later (LL) reward. In these theoretical curves, choice is calculated according to the multiplicative hyperbolic model of Ho *et al.* (1999). Subjects calculate value according to the formula $V = \frac{1}{1+K \cdot d} \times \frac{V_{\max}}{1+Q/q}$ where V is overall value, d is delay, q is quantity, $V_{\max}$ is the maximum possible value that a reward can have, K is a temporal (delay) discounting parameter and Q is a quantity (magnitude) discounting parameter. Subjects then allocate preference (e.g. relative number of choices or relative response rate) in proportion to the values of the two alternatives, i.e. $\frac{pref_A}{pref_A + pref_B} = \frac{V_A}{V_A + V_B}$. **(a)** These curves show the effect of altering K while holding Q constant. "Impulsive" subjects—those who are more likely to choose the SS reward over the LL reward—have a higher value of K than "self-controlled" subjects; their Q parameters do not differ. **(b)** These curves show the effect of altering Q while holding K constant. "Impulsive" subjects have a lower value of Q than "self-controlled" subjects; their K values do not differ.

Models such as the multiplicative hyperbolic discounting model of Ho *et al.* (1999) have been derived based on behavioural techniques allowing magnitude and delay discounting parameters to be determined independently. Unfortunately, the behavioural technique used by Cardinal *et al.* (2001) cannot be analysed using this model. For example, sham subjects' preferences approached 100% choice of the large reinforcer at zero delays (Figure 18a), whereas in the model of Ho *et al.* (1999), relative preference between a one-pellet and a four-pellet reinforcer cannot exceed 80%. The behavioural result comes as no surprise, for it is the well-known phenomenon of maximization on discrete-trial schedules (see Mackintosh, 1974, pp. 190–195), but it implies that behaviour in this task cannot be quantified according to the hyperbolic discounting model.

However, hypersensitivity to the effects of delay appears somewhat more likely than alterations in reward magnitude processing as an explanation for the effects of AcbC lesions. As discussed above, AcbC lesions also reduced preference for the large reinforcer somewhat at zero delay (Cardinal *et al.*, 2001), but this was probably due to a task artefact, namely within-session generalization from trials in which delays were present (Cardinal *et al.*, 2000). When delays were consistently absent, AcbC-lesioned rats preferred the larger reward to the smaller (Cardinal *et al.*, 2001; 2003b). Further evidence supports the assertion that AcbC-lesioned rats can discriminate large from small rewards. Excitotoxic lesions of the whole Acb do not prevent rats from detecting changes in reward value, induced either by altering the concentration of a sucrose reward or by changing the deprivational state of the subject (Balleine & Killcross, 1994). Such lesions also do not impair rats' ability to respond faster when environmental cues predict the availability of larger rewards (Brown & Bowman, 1995), and nor does inactivation of the Acb with local anaesthetic

or blockade of α-amino-3-hydroxy-5-methyl-4-isoxazolpropionate (AMPA) glutamate receptors in the Acb (Giertler *et al.*, 2004); the effects of intra-Acb NMDA receptor antagonists have varied (Hauber *et al.*, 2000; Giertler *et al.*, 2003). AcbC-lesioned rats can still discriminate large from small rewards (Cardinal *et al.*, 2003b; 2004). Similarly, DA depletion of the Acb does not affect the ability to discriminate large from small reinforcers (Salamone *et al.*, 1994; Cousins *et al.*, 1996; Salamone *et al.*, 2001). However, these studies do not address the question of whether AcbC lesions alter the *quantitative* assessment of reward magnitude—e.g. whether such lesions alter Q in the model of Ho *et al.* (1999). Systemic DA antagonists do not affect the perceived quantity of food as assessed in a psychophysical procedure (Martin-Iverson *et al.*, 1987), but this is uninformative as to the role of the AcbC specifically.

The observation that AcbC lesions reduce preference for delayed, certain rewards (Pothuizen *et al.*, 2005) as well as delayed, large rewards (Cardinal *et al.*, 2001), is also consistent with the hypothesis that AcbC-lesioned animals have an impaired tolerance for delays, and that the effects are not due simply to effects on reward magnitude processing—though at present, the role of the AcbC in choosing uncertain reinforcement is also unclear (discussed further below). Acb lesions have also produced delay-dependent impairments in a delayed-matching-to-position task (Reading & Dunnett, 1991).

1.11.1.3 The matching law and reinforcer magnitude assessment

Semi-quantitative assessment of delay and magnitude discounting may be possible in delay-of-reinforcement choice tasks using indifference-point methodology (Ho *et al.*, 1999). Alternatively, relative preference for two reinforcers may be inferred from the distribution of responses on concurrent variable interval (VI) schedules of reinforcement. This literature stems from the discovery by Herrnstein (1961; 1970) of the "matching law". Herrnstein (1961) trained pigeons to respond on two concurrent VI schedules, and varied the relative availability of reinforcement on the two schedules while holding the overall reinforcement rate constant. He observed that the proportion of the total behaviour allocated to each response key approximately matched the proportion of reinforcers allocated to that key. This defines the matching law:

$$\frac{R_1}{R_1 + R_2} = \frac{r_1}{r_1 + r_2}$$

where R represents the behavioural response rate for each alternative, and r the reinforcement. Herrnstein (1970) extended this relationship to take account of more than two alternatives, particularly including "unmeasured" activities the animal may engage in, and derived a "general principle of response output" (Herrnstein, 1970, p. 256):

$$R_1 = \frac{kr_1}{r_1 + r_e}$$

where R_1 is the rate of the response being measured, r_1 is the quantity of reinforcement for that response, r_e is the reinforcement for all other responses, and k is a parameter determining the maximum response rate. Although there are situations where the matching law is not useful—in particular, ratio schedules, where the distribution of reinforcement necessarily *follows* the distribution of responding—a large body of work has sought to define the effects of varying parameters of reinforcement (such as rate, probability, delay, and magnitude) based on this principle (see de Villiers & Herrnstein, 1976).

This technique has not been without problems; in many circumstances, subjects have been found to "overmatch" (exhibit preferences that are exaggerated relative to the predictions of the matching law) or "undermatch" (exhibit reduced relative preferences). This has led to further development of the mathematical models (Baum, 1974; Baum, 1979), though it has been argued in some cases that this approach is circular (Rachlin, 1971). Maximum response rates (k in the equation above) have been shown to vary with the kind of reinforcement used (Belke, 1998), violating an assumption of Herrnstein's law. Nevertheless, the matching law and its extensions do a good job of describing the relationship between reinforcement rate and behaviour on concurrent VI and concurrent-chain schedules (Williams, 1994).

This allows for the possibility of assessing the effects of AcbC lesions upon reinforcer magnitude perception semi-quantitatively (Cardinal, 2001). Used with identical schedules delivering large and small rewards, the matching technique could be used to assess whether or not AcbC-lesioned rats exhibited relative indifference ("undermatching" compared to shams) between the reinforcers used by Cardinal *et al.* (2001). This would provide evidence for reduced reinforcer magnitude discrimination following AcbC lesions, or for an abnormality of the matching process itself, while normal performance (or overmatching) would make this explanation less likely and therefore support the view that AcbC lesions produce a steeper delay-of-reinforcement gradient. As yet, published data do not allow this question to be answered.

1.11.1.4 Learning with delayed reinforcement

If AcbC lesions do indeed induce hypersensitivity to delays of reinforcement, then the effects of AcbC lesions might also extend to *learning* with delayed reinforcement, as well as choice involving delayed reinforcers. In order to learn which actions are the correct ones that eventually lead to reinforcement, some mechanism must "bridge" the delay between action and outcome. Action–outcome delays impair instrumental learning in normal animals to some degree (Grice, 1948; Lattal & Gleeson, 1990; Dickinson *et al.*, 1992). If the AcbC is critical for learning with delayed reinforcement, then AcbC lesions should induce a delay-dependent impairment in free-operant learning with action–outcome delays. This prediction has not yet been tested.

1.11.1.5 Choice involving uncertain reward

Correlational studies have also suggested that the Acb may also be involved in the processing of uncertain or probabilistic reinforcement. DA neurons that innervate the Acb may fire in a manner related to reward probability (Fiorillo *et al.*, 2003; Fiorillo *et al.*, 2005; Niv *et al.*, 2005; Tobler *et al.*, 2005) and the midbrain, the site of the cell bodies of these neurons, responds to stimulus uncertainty in humans (Aron *et al.*, 2004). A greater blood-oxygen-level-dependent (BOLD) response is observed in the human Acb during the selection of high-reward/high-risk options, compared to low-reward/low-risk outcomes, in a task where the risk is of not winning (Ernst *et al.*, 2004), with similar activation to high-reward/high-risk option selection in a task where the risk is of losing (Matthews *et al.*, 2004); this latter activation was correlated with personality measures of harm avoidance. Likewise, an increase in Acb activation (BOLD signal) preceded risk-taking decisions in a financial game with human subjects (Kuhnen & Knutson, 2005). However, to date no studies have examined the contribution of the AcbC to choice involving reward uncertainty. In a recent interventional study, AcbC lesions reduced preference for delayed, certain rewards (Pothuizen *et al.*, 2005), but this does not specifically address the contribution of the AcbC to choosing rewards based on their certainty, particularly as there is evidence suggesting that AcbC lesions impair the processing of delayed reinforcement (Cardinal *et al.*, 2001; Pothuizen *et al.*, 2005).

1.11.1.6 Relationship to neuromodulator function

The Acb is innervated by the a number of neuromodulator systems, including 5-HT (see Halliday *et al.*, 1995) and DA (Ungerstedt, 1971; Fallon & Loughlin, 1995). The DA projection to the Acb is prominent, but although systemic D_2-type DA receptor antagonists can induce impulsive choice involving delayed reinforcement (Wade *et al.*, 2000), this effect may not depend critically on DA in the Acb. Intra-Acb D_1 and D_2 receptor antagonists do not affect rats' ability to wait for reward in a cued progressive delay task (Wakabayashi *et al.*, 2004), and DA depletion of the Acb using 6-OHDA appears not to affect delay discounting directly, though it modifies the effect of systemic 5-HT_{1A} receptor agonists on choice between SS and LL rewards (Winstanley *et al.*, 2005b). The Acb does not receive a substantial NA innervation (Aston-Jones *et al.*, 1995).

1.11.2 Nucleus accumbens shell (AcbSh)

In contrast to the effects of AcbC lesions on choice between delayed, certain and immediate, uncertain rewards, AcbSh lesions have been shown to have effects neither on this task nor on DRL efficiency (Pothuizen *et al.*, 2005). The AcbSh responds to a variety of USs (Bassareo & Di Chiara, 1999; Ito *et al.*, 2000) and has a role in the hedonic assessment of rewards (Berridge, 2000; Kelley & Berridge, 2002). It plays a role in latent inhibition (Pothuizen *et al.*, 2005; 2006), and influences unlearned behaviours including feeding (Kelley & Swanson, 1997; Stratford & Kelley, 1997; Basso & Kelley, 1999; Kelley, 1999) and locomotion (Swanson *et al.*, 1997; Parkinson *et al.*, 1999a). The AcbSh has also been shown to be abnormal in animal models of ADHD (Papa *et al.*, 1996; Carey *et al.*, 1998; Papa *et al.*, 1998; Sadile, 2000). However, these results suggest it does not contribute to choice involving delayed or uncertain rewards (Pothuizen *et al.*, 2005).

1.11.3 Anterior cingulate cortex (ACC)

Excitotoxic lesions of the ACC (Figure 16) have no effect on choice between SS and LL rewards in rats (Cardinal *et al.*, 2001) (Figure 17), indicating that the ACC is not required for normal choice of delayed reinforcement. These results suggest that ACC dysfunction is not an important contributor to impulsive choice, despite the involvement of the ACC in reward-related learning (Bussey *et al.*, 1997a; Bussey *et al.*, 1997b; Parkinson *et al.*, 2000c; Cardinal *et al.*, 2003a) and findings of ACC abnormalities in ADHD (Bush *et al.*, 1999; Rubia *et al.*, 1999). However, ACC lesions do impair choice between small/sooner/low-effort and large/later/high-effort alternatives, reducing preference for the high-effort option (Walton *et al.*, 2002; 2003), indicating that the ACC is involved in promoting the selection of effortful alternatives. The DA innervation of the ACC does not appear important for this function (Walton *et al.*, 2005).

However, ACC lesions can make rats motorically impulsive, with simple disinhibition or "execution" impulsivity. ACC-lesioned rats have been found to over-respond to unrewarded stimuli (Bussey *et al.*, 1997a; Parkinson *et al.*, 2000c) and to respond prematurely in situations where they are required to wait (Muir *et al.*, 1996). They also exhibit discriminative deficits in Pavlovian conditioning tasks (Cardinal *et al.*, 2003a), though the full range of functions associated with the ACC (including error detection, attentional control, and mood; see e.g. Devinsky *et al.*, 1995; Volkow *et al.*, 1997; Maas *et al.*, 1998; Childress *et al.*, 1999; Bush *et al.*, 2000; Garavan *et al.*, 2000; Paus, 2001; Cardinal *et al.*, 2002a; Vogt, 2005) is beyond the scope of this thesis.

The contribution of the ACC to probabilistic choice is less clear. In both humans and rhesus monkeys, the ACC responds to anticipated gain in tasks in which rewards of different magnitudes are available with

varying probabilities. In the rhesus monkey, the ACC responds to some combination of reward size and reward probability (Amiez *et al.*, 2005a) and deactivation of the ACC impairs such choices (Amiez *et al.*, 2005b), but human studies would suggest that the ACC responds to the magnitude rather than the probability of expected gains (Rogers *et al.*, 2004b). However, a nearby region of human medial PFC has been observed to respond to reward probability (Knutson *et al.*, 2005).

1.11.4 Prelimbic (PrL) and infralimbic (IL) cortex

The mPFC projects to the AcbC, is involved in reward-related learning (e.g. Balleine & Dickinson, 1998; Richardson & Gratton, 1998; Bechara *et al.*, 1999; Tzschentke, 2000), receives DA and 5-HT input (see Fallon & Loughlin, 1995; Halliday *et al.*, 1995), and has been observed to be abnormal in ADHD (Ernst *et al.*, 1998; Rubia *et al.*, 1999). However, lesions of the rat mPFC, primarily PrL and infralimbic (IL) cortex (Figure 16), had no delay-specific effects on choice between large/delayed and small/immediate rewards (Cardinal *et al.*, 2001) (Figure 17); the effects observed appeared to be task-specific, related to an insensitivity to the contingencies or stimuli present in the task, perhaps as a result of a loss of temporal discriminative stimulus control (Cardinal *et al.*, 2003b). It is important to note that PrL may have more functional homology to the primate dorsolateral PFC than to regions that are medial within human PFC (Uylings *et al.*, 2003). Aspirative lesions of the mPFC have previously been shown to induce a deficit in timing ability in rats (Dietrich & Allen, 1998), with impaired temporal discrimination in the peak procedure, an operant task that assesses the ability to time a stimulus (Catania, 1970; Roberts, 1981). Consistent with the view that mPFC lesions did not affect the basic process of choosing between reinforcers of different value, combined PrL/IL lesions did not affect choice between small/low-effort and large/high-effort alternatives in the task of Walton *et al*. (2003).

1.11.5 Orbitofrontal cortex (OFC)

The OFC is a region of the PFC that projects to the AcbC and is strongly implicated in the assessment of reward value. Mobini *et al*. (2002) recently found that lesions encompassing the OFC induced impulsive choice in a discrete-trial SS/LL reward choice task very similar to that described above (Figure 20). As before, results from this simple form of task do not indicate whether the impulsive choice was as a result of altered sensitivity to reinforcer magnitude or delay. Although these lesions damaged prelimbic cortex (PrL) in addition to the OFC (Mobini *et al.*, 2002), the hypothesis that OFC damage was responsible for the behavioural effect is strengthened by the finding that mPFC lesions encompassing PrL do not induce impulsive choice (Cardinal *et al.*, 2001). In contrast, Winstanley *et al*. (2004b) recently found that OFC lesions induced the opposite effect—better self-control than shams (Figure 21)—in exactly the task described above (Figure 15, p. 38). One possible reason for this discrepancy is that subjects in Winstanley *et al*.'s study were trained before the OFC was destroyed and retested postoperatively, while Mobini *et al*. trained and tested postoperatively. Another is that Mobini *et al*. (2002) offered rats a choice between a one-pellet immediate reinforcer and a two-pellet delayed reinforcer, whereas Winstanley *et al*. (2004b) used a one-pellet immediate reinforcer and a four-pellet delayed reinforcer. As discussed above, differences in subjects' sensitivity to either the delay or the magnitude of reinforcement can play a role in determining preference in this task (Ho *et al.*, 1999; Mobini *et al.*, 2002; Cardinal *et al.*, 2003b) and it may be that OFC lesions affect both, increasing both the delay discounting parameter K and the magnitude discounting parameter Q (Mobini *et al.*, 2002). An increase in K would imply steeper delay discounting; an increase in Q would imply an increase in sensitivity to the ratio of the magnitudes of the two reinforcers, and could mask (or potentially reverse) the increase in impulsivity produced by the increase in K.

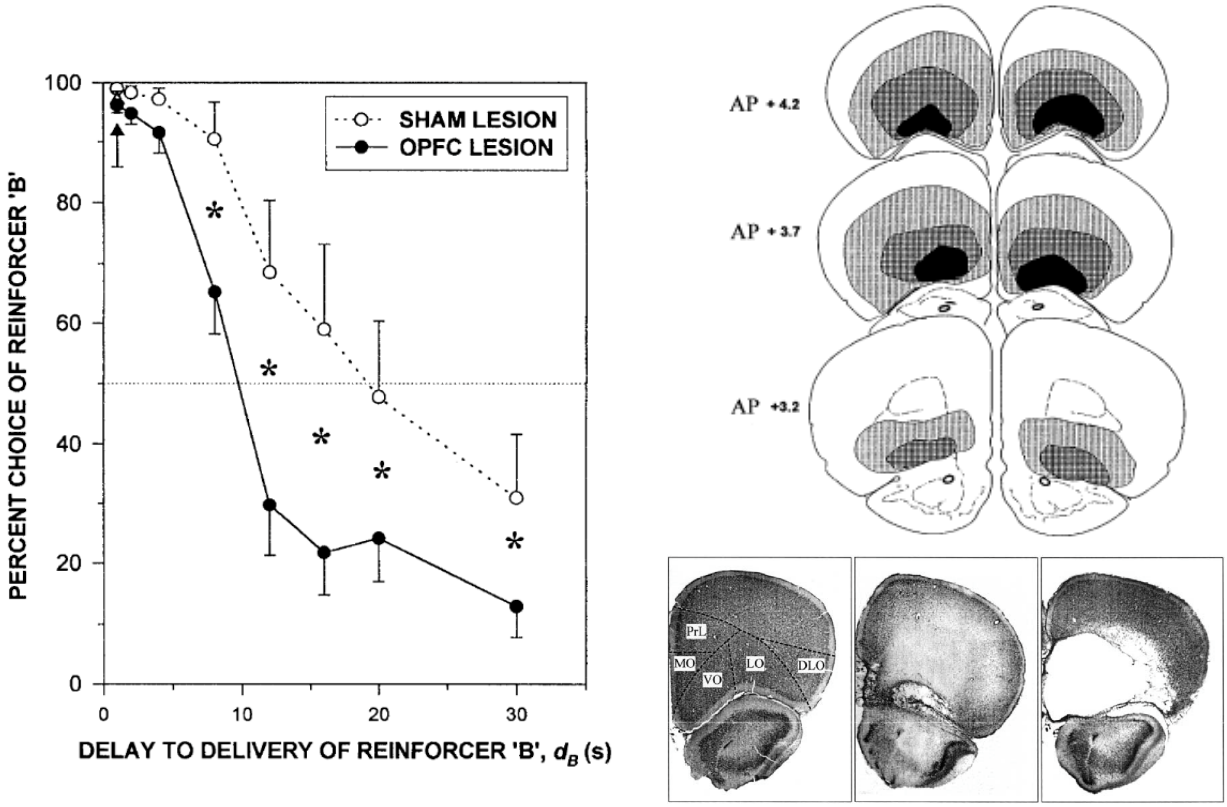

Figure 20: Choice between immediate, small and large, delayed rewards in rats with lesions of the orbitofrontal cortex (OFC)

In this study, rats received excitotoxic lesions of the OFC or sham lesions before being trained on a delayed reinforcement choice task similar to that described earlier, but involving choice between a one-pellet small reward and a two-pellet large reward. Delays to the large reinforcer varied across, rather than within, sessions. **(Left)** OFC-lesioned rats were more impulsive than shams, choosing the larger/later reward less often. **(Right)** Lesion schematics and representative photomicrographs. From Mobini *et al.* (2002).

There is direct support for this hypothesis: OFC lesions appear to increase *K*, the rate of delay discounting, as well as increasing the magnitude sensitivity parameter *Q* (Kheramin *et al.*, 2002; Kheramin *et al.*, 2003). The same effect of increases in both *K* and *Q* has been observed with DA-depleting OFC lesions (Kheramin *et al.*, 2004). This emphasizes the necessity for quantitative analysis of delay and magnitude sensitivity (Ho *et al.*, 1999) or the use of multiple, very different paradigms to provide independent measurements of sensitivity to delay and magnitude (Cardinal *et al.*, 2003b). It also reminds us of an important clinical point: faced with steep delay discounting in a task involving choice between SS and LL rewards, increasing the ratio of the large to the small reward may ameliorate the impulsivity.

As discussed above, it has been suggested that hyperbolic discounting is explicable as the overall effect of two or more different systems, such as an explicit (declarative) system that exhibits minimal or exponential discounting, plus phenomena that make rewards more salient and promote their choice when they are immediately available. Recently, such a two-factor model was used in the analysis of a functional magnetic resonance imaging (fMRI) study of choice involving rewards differing in magnitude and delays, with delays ranging from less than a day to 6 weeks (McClure *et al.*, 2004): lateral prefrontal and intraparietal cortical regions were activated independently of the delay, and were suggested to be part of a system that evaluates both immediate and delayed rewards according to a "rational" (meaning non-hyperbolic) temporal discounting system, while limbic regions including the ventral striatum and medial OFC were preferentially activated by the relatively immediate rewards, and were suggested to be part of a

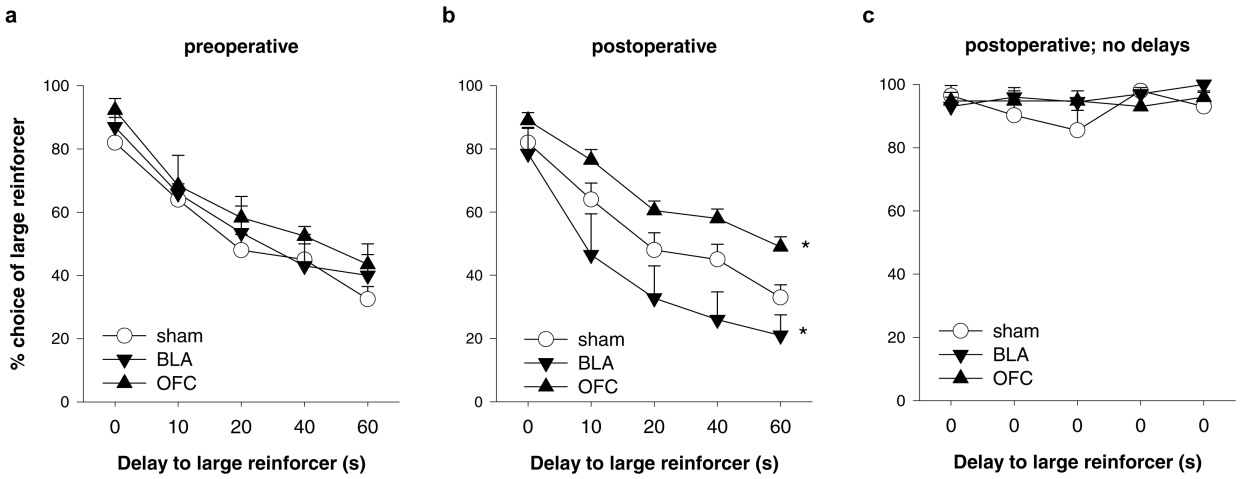

Figure 21: Choice between immediate, small and large, delayed rewards in rats with lesions of the basolateral amygdala (BLA) or OFC

Rats were trained on the delayed reinforcement choice task (involving choice between a one-pellet immediate reinforcer and a four-pellet large reinforcer, and within-session changes in the delay to the large reinforcer; Figure 15) before matched groups of subjects **(a)** received excitotoxic lesions of the BLA, or the OFC, or sham lesions; **(b)** shows their postoperative performance, in which BLA-lesioned subjects were more impulsive (more likely to choose a single immediate pellet over a four-pellet delayed reward) and OFC-lesioned subjects were less impulsive than sham-operated controls (* $p < .05$, difference from shams). **(c)** When all delays were removed from the task, BLA- and OFC-lesioned subjects chose identically to shams, preferring the four-pellet reward to the one-pellet reward. Redrawn from Winstanley *et al.* (2004b).

system that promotes the choice of imminent rewards without consideration of delayed alternatives. These limbic regions were more likely to be activated than the "delay-independent" areas on trials in which an earlier reward was chosen. This would sit neatly with studies showing that OFC lesions reduce impulsive choice (Winstanley *et al.*, 2004b); however, it does not square so easily with rodent evidence showing that destruction of the AcbC (a major part of the ventral striatum) or the OFC enhances delay discounting, meaning that delayed alternatives are less likely to be chosen (Cardinal *et al.*, 2001; Kheramin *et al.*, 2002; Mobini *et al.*, 2002; Kheramin *et al.*, 2003).

The PFC, which projects heavily to the AcbC (Brog *et al.*, 1993), is also involved in decision making under conditions of uncertainty. Humans with OFC or ventromedial PFC damage are impaired in the Iowa gambling task (Bechara *et al.*, 1994; 1996; 1997), in which subjects must learn to differentiate between low-reward, low-risk card decks that yield a net positive outcome and high-reward, high-risk decks that yield a net negative outcome, though the precise locus and nature of the deficit seen in this task is debated (Manes *et al.*, 2002; Clark *et al.*, 2003; Fellows & Farah, 2005). OFC neurons respond to reward expectancy (see Hikosaka & Watanabe, 2000). Choice between small, likely rewards and large, unlikely rewards increases blood flow and BOLD signal in orbital and inferior PFC (Rogers *et al.*, 1999b; Ernst *et al.*, 2004; Rogers *et al.*, 2004b), and OFC damage also impairs performance of a task requiring human subjects to choose between two possible outcomes and to bet on their choice, with lesioned subjects deciding slowly and failing to choose the optimal, most likely outcome (Rogers *et al.*, 1999a). Excitotoxic lesions of the OFC make rats less likely than sham-operated controls to choose a large, uncertain reward over a small, certain reward (Mobini *et al.*, 2002); OFC-lesioned rats had lower indifference odds (higher indifference probabilities; steeper uncertainty discounting) and exhibited risk-averse choice. As discussed above, there is direct evidence that excitotoxic OFC lesions and OFC DA depletion do alter sensitivity to the relative magnitudes of the two rewards (Kheramin *et al.*, 2004; Kheramin *et al.*, 2005), but the effect

of steepening uncertainty discounting (of increasing the odds discounting parameter *H*) is present in addition to the effects on reinforcer magnitude sensitivity (Kheramin *et al.*, 2003).

A recent fMRI study also examined human regional cerebral BOLD responses to decision making in a task that explicitly distinguished "ambiguity" from "risk" (Hsu *et al.*, 2005). "Ambiguity" referred to the situation in which the probability of a successful outcome was unknown—such as having to bet on whether the next card drawn from a twenty-card deck containing red or blue cards would be red or blue, with no further information. "Risk" referred to the situation in which the probability of a successful outcome was known, but not zero or one—such as having to bet on the colour of the next card drawn from a deck containing ten red and ten blue cards. This task produces behaviour that is economically self-contradictory. Subjects prefer to bet on red from the risky deck than on red from the ambiguous deck, but also prefer to bet on blue from the risky deck than on blue from the ambiguous deck (Becker & Brownson, 1964; MacCrimmon, 1968). These preferences are mutually inconsistent under simple probability theory—termed the Ellsberg paradox (Ellsberg, 1961)—and imply an inherent aversion to ambiguity. In these tasks, the OFC, amygdala, and dorsomedial PFC were more active under conditions of ambiguity than risk, while dorsal striatal activity showed the opposite pattern (risk > ambiguity). While normal humans exhibited aversion to ambiguity, and also aversion to risk (in other tasks in which the card proportion was varied), humans with OFC lesions were averse neither to risk nor ambiguity—behaviourally abnormal, but consistent with expected utility theory (Hsu *et al.*, 2005).

1.11.6 Insula

A further cortical region that may be involved in decisions involving uncertainty is the insula, or insular cortex. Anterior insula activation has been observed to precede risk-averse choice in humans (Kuhnen & Knutson, 2005), in a task in which Acb activation preceded risk-prone choice. The authors suggest that in tasks such as these, the Acb represents predictions of gain (Knutson *et al.*, 2001), while the insula represents predictions of loss (see also Paulus *et al.*, 2003); activation in both structures is related to personality measures of harm avoidance (Paulus *et al.*, 2003; Matthews *et al.*, 2004).

1.11.7 Basolateral amygdala (BLA)

The basolateral amygdala (BLA) also projects to the AcbC, and has extensive reciprocal connections with the OFC. Excitotoxic lesions of the BLA promote impulsive choice in a task involving choice between an immediate one-pellet reward and a delayed four-pellet reward (Winstanley *et al.*, 2004b) (Figure 21), similar to the effects of AcbC lesions (Cardinal *et al.*, 2001) but opposite to those of OFC lesions in the same task (Winstanley *et al.*, 2004b) (Figure 21). Although this study is notable for finding opposite effects of BLA and OFC lesions, which is unusual (see also Izquierdo & Murray, 2005), the explanation for this effect is unclear. One obvious possibility, given the effects of OFC lesions to increase both the delay discounting parameter *K* and the magnitude sensitivity parameter *Q* (in the model of Ho *et al.*, 1999), is that BLA lesions and AcbC lesions simply increase *K* without affecting *Q* (cf. Figure 19, p. 41). There is indirect evidence for this in the case of AcbC lesions, discussed above; for the BLA, this hypothesis remains untested. Some studies have demonstrated deficits following amygdala inactivation when reward size is suddenly changed (Salinas *et al.*, 1993; Coleman-Mesches *et al.*, 1996; Salinas & McGaugh, 1996; Salinas *et al.*, 1997; Liao & Chuang, 2003), though changing the size of a reward for performing the same task has obvious emotional significance and the amygdala is well known to be involved in affective representation (see Aggleton, 2000; Cardinal *et al.*, 2002a). One study has found deficits in *memory* for reinforcer magnitude following amygdala lesions, even if this was not a primary

deficit in reinforcer magnitude discrimination (Kesner & Williams, 1995). None of these bear directly on the question of whether relative reinforcer magnitude discrimination (as measured by *Q*) is altered by BLA lesions.

A recent study has also suggested the involvement of the BLA in promoting the selection of effortful alternatives. Floresco & Ghods-Sharifi (2006) showed that BLA inactivation with the local anaesthetic bupivacaine impaired rats' ability to choose a large/high-effort alternative over a small/low-effort alternative. This is much like the effect of ACC lesions discussed above (Walton *et al.*, 2002; 2003), and indeed, a reversible BLA–ACC disconnection lesion also impaired selection of large/high-effort alternatives (Floresco & Ghods-Sharifi, 2006), suggesting that direct information transfer between the BLA and the ACC is important in this task.

1.11.8 Subthalamic nucleus (STN)

The subthalamic nucleus (STN) is a component of the basal ganglia that receives projections both from the globus pallidus (pallidum) and the cerebral cortex (Alexander & Crutcher, 1990; Hamani *et al.*, 2004) and projects to basal ganglia relay structures (including the globus pallidus, the rodent homologue of the external part of the primate globus pallidus) and output structures of the basal ganglia, including the entopeduncular nucleus and the substantia nigra pars reticulata (Heimer *et al.*, 1995; Hamani *et al.*, 2004), which project on to thalamus and thence to cortex. Lesions of the STN decreased impulsive choice in a task involving choice of a single immediate food pellet or four pellets delivered after a delay (Winstanley *et al.*, 2005a), a task in which OFC lesions had the same effect (Winstanley *et al.*, 2004b). STN lesions also impaired autoshaping (Winstanley *et al.*, 2005a), meaning locomotor approach to appetitive Pavlovian CSs (Brown & Jenkins, 1968; Williams & Williams, 1969). However, this is unlikely to explain the effect of STN lesions to promote choice of LL rewards—not least because AcbC lesions also impair autoshaping (Parkinson *et al.*, 2000c; Cardinal *et al.*, 2002b) but reduce choice of LL rewards (Cardinal *et al.*, 2001), while ACC lesions impair autoshaping (Bussey *et al.*, 1997a; Parkinson *et al.*, 2000c; Cardinal *et al.*, 2003a) but do not affect choice between SS/LL rewards (Cardinal *et al.*, 2001), but more simply because there was no explicit CS in this task differentially associated with the two rewards, and approach to which would promote choice of the SS reward. Furthermore, STN lesions tend to increase premature responding, often thought of as an index of motor impulsivity (Baunez & Robbins, 1997; Baunez *et al.*, 2001). It is not known whether STN lesions affect reward magnitude discrimination or uncertainty discounting.

1.11.9 Hippocampus (H)

Finally, a role of the hippocampus in learning with delayed reinforcement might be suspected. As discussed earlier, contextual conditioning is important in learning with delays, and there is good evidence that the hippocampus contributes to the representation of context. Context-specific representations develop in the hippocampus (Smith & Mizumori, 2006), and lesions of the hippocampal formation (H) have been shown to impair Pavlovian conditioning to a contextual CS, but not to a discrete CS, in rats (Hirsh, 1974; Selden *et al.*, 1991; Kim & Fanselow, 1992; Phillips & LeDoux, 1992; Honey & Good, 1993; Jarrard, 1993; Kim *et al.*, 1993; Phillips & LeDoux, 1994; Phillips & LeDoux, 1995; Chen *et al.*, 1996; Maren & Fanselow, 1997; Anagnostaras *et al.*, 1999; Rudy *et al.*, 2002), at least for some processes involving contextual representation (Good & Honey, 1991; Holland & Bouton, 1999; Good, 2002). In some cases, discrete CS conditioning has even been enhanced (e.g. Ito *et al.*, 2005). Since context–outcome associations are thought to hinder instrumental learning with delayed reinforcement through contextual

competition (Dickinson *et al.*, 1992; Dickinson & Balleine, 1994), it follows that if H lesions impair the formation of associations involving the context, such lesions might reduce contextual competition and hence *facilitate* instrumental conditioning when there is an action–outcome delay.

Despite this clear prediction, the contribution of the hippocampus to learning with delayed reinforcement, or to self-controlled choice, has not previously been investigated in detail. One previous study found that aspirative lesions of the dorsal hippocampus did not affect appetitive instrumental conditioning with delayed reinforcement (Port *et al.*, 1993), but this study was poorly designed to address this question in a number of ways; amongst its flaws, the study used aspirative rather than excitotoxic lesions, used a task in which alterations in response rates affected the instrumental contingency, and tested subject at a single delay with no zero-delay control condition.

The only study to date to address the influence of the hippocampus on choice involving delayed or uncertain reward was that of Rawlins *et al.* (1985), who examined choice between certain and uncertain rewards. Normal rats preferred immediate certain reward to immediate uncertain reward, and also preferred delayed certain reward to immediate uncertain reward; however, rats with hippocampal or medial septal lesions were less tolerant of the delay (or more tolerant of the uncertainty), preferring immediate uncertain reward to delayed certain reward. However, this study does not answer the question of whether hippocampal lesions affect the processing of reward delay or reward uncertainty specifically.

1.12 OUTLINE OF EXPERIMENTAL WORK IN THIS THESIS

This thesis has three principal objectives: first, to establish whether the role of the AcbC in choosing large, delayed rewards reflects an underlying deficit in the processing of reward delay or of reward magnitude; second, to investigate the role of the AcbC in decisions involving uncertain reward; and third, to establish the role of the hippocampus in the processing of delayed reward. Chapter 2 will examine the role of the AcbC in free-operant learning with delayed reward, performance of a previously learned free-operant response for delayed reward, and the quantitative allocation of behaviour to match obtained reward magnitudes. Chapter 3 will examine the role of the hippocampus in learning with delayed reward, performance of a previously learned free-operant response for delayed reward, and choice between SS and LL reward alternatives. Chapter 4 will return to the AcbC, examining its role in choice between small/certain and large/unlikely rewards.