BEHAVIOURAL AND NEUROCHEMICAL ANALYSIS OF STIMULUS-RESPONSE HABITS IN RATS

A thesis submitted for the degree of Doctor of Philosophy at the University of Cardiff

Andrew J D Nelson B.A. B.Sc.

2006

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ..... (candidate) Date 10/11/06

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed ..... (candidate) Date 10/11/06

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.
Other sources ~~are~~ acknowledged by explicit references.

Signed ..... (candidate) Date 10/11/06

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate) Date 10/11/06

STATEMENT 4 - BAR ON ACCESS APPROVED

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access** approved by the Graduate Development Committee.

Signed ..... (candidate) Date 10/11/06

CONTENTS

PUBLICATIONS

Nelson, A., & Killcross, S. (2005). Repeated amphetamine pre-treatment renders actions insensitive to outcome devaluation. *Behavioural Pharmacology*, 16, S57-S57 Suppl. 1.

Nelson, A., & Killcross, S. (2006). Amphetamine exposure enhances habit formation. *Journal of Neuroscience*, 26 (14), 3805-3812.

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor Prof. Simon Killcross whose guidance and enthusiastic support made the completion of this project possible. I would like to thank Dr. Jo Haddon who assisted me greatly. I would also like to acknowledge Dr. Dave George for help with writing (and debugging) Med PC programmes. Thanks are also due to Lyn Hill and Kerry Housler who provided animal care.

This work was supported by a BBSRC studentship.

SUMMARY

Considerable evidence has accumulated demonstrating that instrumental actions in rats can be controlled by two dissociable associative structures. During the early stages of training, responding is guided by action-outcome (A-O) associations that require both a representation of the outcome and knowledge of the instrumental contingency. However after more extended training, behaviour comes to be controlled by stimulus-response (S-R) habits that are no longer goal-directed. Despite the behavioural evidence supporting this dissociation, the psychological and neurochemical mechanisms underpinning this transition are poorly understood. To address this first issue, we compared the sensitivity to outcome devaluation of instrumental responses that were moderately or extensively trained on either interval or ratio schedules. It was found that S-R habits developed as animals achieved stable rates of responding. This was further supported by demonstrations that when well-established performance is disrupted by changes in the schedule of reinforcement or reward magnitude, or where there is no consistent relationship between behaviour and reward delivery animals appear sensitive to goal-value despite extended training. It is suggested that S-R habits develop as the molar correlation between behaviour and reward becomes well-predicted. Moreover, the work presented here also sought to elucidate further the neurochemical processes involved in the transition from action to habit. These experiments demonstrated that pre-training exposure to amphetamine leads to the early and excessive dominance of S-R processes. This provides the first direct evidence that dopamine transmission is critical to the

development of S-R habits. Further experiments explored the neuropharmacological specificity of this effect and found a dissociation at the level of the receptor subtype: amphetamine enhancement of S-R learning is reversed by D_1, but enhanced by D_2, receptor antagonists. Finally, the results are discussed in terms of optimality and certainty-based models of instrumental performance with reference to both phasic and tonic dopamine activity.

CHAPTER 1

INTRODUCTION

Habits and actions: the dual control of behaviour

In his Principles of Psychology of 1890, James recognised the pervasion of habits in both human and non-human behaviour. 'When we look at living creatures from an outward point of view, one of the first things that strikes us is that they are bundles of habits' (pp. 104). While it would seem evident that much behaviour is characterised by voluntary, goal-directed actions, outwardly similar behaviours can equally be performed automatically as habits. For example, the simple action of switching a light on might be the result of a need for illumination while reading a book, but may also be an automatic response to entering a dark room. The former is an example of a goal-directed action guided by outcome expectancy (the explicit desire and need for light) as well as declarative knowledge about the relationship between flicking a switch and light; the latter is a habitual response to a stimulus in the environment (a light switch in the context of a dark room) and is without regard for consequences. Similarly, after initial acquisition, seemingly complex behaviours such as driving can be executed effortlessly and automatically. Habits can also lead to the ubiquitous and rather frustrating phenomenon of actions slips (the performance of actions that are unattended). Reason (1979) reports the case of a participant in a diary study 'I meant to get my car out, but as I passed through the back porch on my way to the garage I stopped to put on my Wellington boots and gardening jacket as if to work in the garden.' Thus, unlike goal-

directed actions which are controlled by their consequences, habits are governed by antecedent stimuli. Complex repertoires of actions can be triggered by the mere presence of a stimulus in the environment. Furthermore, these actions can be performed without effort, regard for the outcome or even conscious awareness. Habits, therefore, represent the behavioural expression or output of non-declarative procedural memory. The ability to perform behaviours as habits endows organisms with obvious biological advantages. Purposeful, goal-directed behaviour places considerable demands on cognitive resources such as attention and working memory. Behaviour that can be elicited and performed automatically, therefore, frees up cognitive resources and allows attention to be directed to the attainment of other goals (Shiffin and Schneider, 1977).

The study of habits is not only of significance in elucidating normal human and non-human cognition, it can also be applied to the field of abnormal psychology. Indeed, habits can become so entrenched that they persist in spite of changed contingencies in the environment. For example, the action of switching a light on when entering a dark room may persist even when the light switch has broken. While this may be an innocuous example of response perseveration, it is this aspect of habit learning, the resistance to extinction and imperviousness to changes in outcome value, which is of relevance to the understanding of certain human psychopathologies. As behaviour comes to be elicited not by its consequences but by its antecedents, responses can become inappropriate and maladaptive. Tolman (1932) recognised that animals can become 'fixated' upon an activity and the notion of 'bad habits' forms part of Folk psychology. Indeed, drug addiction is a pernicious disease which has deleterious consequences for both the

individual concerned and society in general. The transition from acute drug abuse to chronic and compulsive drug addiction can be viewed, inter alia, as the result of aberrant habit learning (Robbins and Everitt, 1999; Everitt, Dickinson and Robbins, 2001; Dickinson, Wood and Smith, 2002; Miles, Everitt and Dickison, 2003; Everitt and Robbins, 2005; Hyman, Malenka and Nestler, 2006). Similarly, many human neuropsychiatric conditions, such as obsessive-compulsive disorder, Tourette's syndrome, autistic spectrum disorder and schizophrenia, are marked or even defined by the production of maladaptive, involuntary and repetitive behaviours (Ridley, 1994; Graybiel and Rauch, 2000; Leckman and Riddle, 2000). Thus advances in understanding both the behavioural and neural substrates of habit learning will prove critical in expounding the psychopathogenesis of these disorders.

Behavioural dissociation of actions and habits in rats

The behavioural substrates of habit learning in animals can be investigated by analysing the associative structure underlying instrumental conditioning. In instrumental conditioning, reward delivery is explicitly contingent on specific responses. This response may be lever pressing, chain pulling, traversing a t-maze or flexing a leg. Whatever the response, the occurrence or avoidance of an unconditioned stimulus (UCS) is contingent upon the animal's behaviour. This contrasts with other forms of learning such as Pavlovian conditioning. In Pavlovian conditioning there is no contingency between the animal's response and reward delivery; rather, reward presentation is

dependent on the presence of cues in the environment that come to elicit conditioned responses such as magazine approach.

Although it would seem intuitive that instrumental conditioning is directed towards the attainment of goals and that animals must therefore encode not only the causal relationship between their behaviour and response delivery but also have a representation of outcome value, most early work on instrumental conditioning was couched in behaviourist terms of stimulus-response (S-R) learning (e.g. Hull, 1943). The central tenet of behaviourism is that all behaviour is elicited by stimuli in the environment. Thorndike's Law of Effect (1911) held that reinforcement serves to strengthen the connection between a stimulus and a response and renders it more likely that a specific response will be emitted by the animal in the presence of that stimulus. Thus according to S-R theories of instrumental conditioning, animals form an association between the stimulus or context in which reinforcement occurs and a specific behavioural response. Such a position denies animals knowledge about the effects of their behaviour and hence renders behaviour inflexible and potentially maladaptive. This contrasts with teleological accounts which posit that an animal's instrumental behaviour is controlled by knowledge of the consequences of its actions and an expectancy of the outcome (e.g. Tolman, 1932). This view sees instrumental learning as forming an association between an action and an outcome (A-O) and thereby endows the animal with greater behavioural flexibility. For most of the twentieth century, theories of instrumental conditioning assumed the centrality of S-R associations in instrumental learning (e.g. Hull, 1943; Spence, 1956; Rescorla and Solomon, 1967; Trapold and Overmier, 1972) or afforded a role to outcome

expectancy and A-O associations in guiding instrumental performance (e.g. Bolles, 1972; Mackintosh, 1974), but without behavioural techniques to test empirically these theories, little real progress could be made in elucidating the associative structure underlying instrumental learning.

However, in the last twenty five years a paradigm shift has occurred that has allowed a dissociation of these two possible accounts of instrumental conditioning. This has been achieved through the application of modern behavioural assays that can measure the extent to which instrumental conditioning is controlled by purposeful goal-directed actions (A-O) or reflexive S-R habits. The first behavioural assay involves changing (often devaluing) the value of the reinforcer so that the outcome is no longer motivationally significant or desirable for the animal. For example, the outcome can be paired with lithium-chloride (LiCl) to induce gastric-malaise in the animal. Alternatively, the animal can be sated (specific satiety) on the instrumental outcome. The animal's propensity to respond for the instrumental outcome can then be indexed in a subsequent extinction test. Significantly, the extinction test is conducted in the absence of reward delivery so that it assays the memory of the associative structure rather than any new learning. Performance in this probe test can be compared to that of animals for which the value of the reinforcer has not been manipulated (non-devalued controls). If an animal's instrumental performance is guided by outcome expectancy and is goal-directed, then its responding in this test should be reduced compared to that of non-devalued controls. This selective depression in response rates can be taken as prima facie evidence that animals are sensitive to the current value of the outcome and are able to use this knowledge to

guide their instrumental performance. Conversely, if animals show no sensitivity to the changed value of the reinforcer and respond at equivalent rates to non-devalued controls, then this suggests that their instrumental performance is impervious to the current value of the goal and hence is controlled by S-R habits. The associative structure of instrumental conditioning can be probed further by degrading the action-outcome contingency. This involves reducing the correlation between an animal's behaviour and reward delivery. For example, on omission schedules food is delivered when the animal withholds responding and pressing the lever actually prevents or delays reward delivery. This procedure has been shown, under certain circumstances, to reduce profoundly instrumental responding (Davis and Bitterman, 1971). Sensitivity to manipulations of the action-outcome contingency provides cogent evidence that animals are able to encode the causal relationship between their behaviour and reward delivery. A failure, on the other hand, to detect suppression in lever press rates in response to the degradation of the action-outcome contingency would be highly suggestive of instrumental performance that is stimulus-bound and habitual.

These behavioural assays, outcome devaluation and degrading the action-outcome contingency, have been used extensively to reveal the associative structure underpinning instrumental learning and the circumstances under which instrumental responding is controlled by goal-directed actions and by S-R habits. Adams (1980) trained rats to press a lever for sucrose pellets and subsequently an aversion to the reinforcer was induced by pairing it with LiCl. This devaluation treatment, however, failed to impact on the animals' propensity to press the lever in an extinction test compared to non-devalued

controls. Nevertheless, in a rewarded reacquisition test the devalued group pressed at lower rates compared to controls. Thus, the finding that lever pressing was independent of the current value of the outcome in the extinction test could not be accounted for by a failure to acquire an aversion to the reinforcer and would suggest that these animals' lever press performance was not goal-directed but habitual. Adams (1980) gave animals 5 days of instrumental conditioning: 1 session on a continuous reinforcement schedule (CRF) and 4, 30-minute sessions on a variable interval 60-seconds schedule (VI). This level of training may have been sufficient to produce instrumental behaviour that was independent of goal-value (Adams and Dickinson, 1981a). Indeed, in two subsequent experiments animals earning 50 reinforcers on a CRF schedule and a further 50 on a variable ratio 9 (VR) schedule were shown to be sensitive to the changed value of the reinforcer after taste aversion training (Adams and Dickinson, 1981b). These results provided empirical support for the position that instrumental responding can be guided by the current value of the reinforcer but that the level or type of training may reduce sensitivity to reward value.

Indeed, it has been a widely held view in both Folk and experimental psychology that extended practice of an instrumental response promotes habit formation (e.g. Kimble and Perlmuter, 1970). To test the hypothesis that sensitivity to outcome value is predicted by the extent of instrumental training, Adams (1982) compared the performance of animals given either limited or extended training in a reinforcer devaluation task. After magazine training, animals either earned 100 or 500 rewards on a CRF schedule. Subsequently, for half the animals an aversion to the reinforcer was induced by LiCl and for the other half

the value of the reinforcer was maintained. In order to obviate any contextual interference on taste aversion training, the LiCl injections were administered after the animals had non-contingent experience of the instrumental outcome in the experimental chambers. Non-devalued controls were placed in the chambers but without exposure to the reinforcer, prior to saline injections. The animals' willingness to press the lever was then assessed in an extinction test. In order to confirm that taste aversion conditioning had proceeded successfully, response rates were measured in a rewarded re-acquisition test. The results of the extinction test demonstrated that instrumental conditioning can be subserved by two distinct and dissociable associative structures. Animals in the limited training condition, averted from the reinforcer, pressed the lever at significantly lower rates compared to the non-devalued controls. Conversely, over-training of the instrumental response resulted in behaviour that was autonomous of the current value of the reinforcer: the devalued group pressed at comparable rates to the non-devalued controls. Significantly in the rewarded reacquisition test all animals in the devalued groups, irrespective of the amount of training received, responded less than the non-devalued controls. Thus, the failure to integrate knowledge about the changed value of the reinforcer after extended training could not be explained in terms of any ineffectiveness of the taste aversion training. Indeed, these results highlight an important difference between extinction and rewarded reacquisition tests. While sensitivity to reward value in an extinction test is evidence of the outcome expectancy of goal-directed actions, lowered responding in reacquisition tests can reflect either the devaluation of A-O associations or the punishment of S-R habits by the presentation of the nausea-inducing reinforcer. Thus, Adams (1982) furnished strong evidence that after limited

training, instrumental performance is sensitive to manipulations in reward value but after more extended training, responding becomes independent of the value of the instrumental outcome. Thus over the course of training, response control is ceded from its consequences to its antecedents.

Adams (1982) identified two factors that may underlie the transition from goal-directed actions to habits. One obvious explanation is that simply repeating a response promotes habit formation. However, the use of a CRF schedule in the above experiment confounded the number of responses emitted and the number of reinforcers earned as possible determinants of the onset of behavioural autonomy. In order to dissociate these two possible factors, after an initial session in which 50 responses were rewarded on a CRF schedule Adams (1982) trained rats on a variable ratio 9 (VR) schedule until the animals had earned a further 50 rewards. Thus animals in this experiment pressed the lever as many times as the 500 condition in the first experiment but earned the same amount of rewards as the 100 condition. Furthermore, Adams (1982) also manipulated the spacing of training and compared sensitivity to outcome devaluation after massed (2 days) or distributed (10 days) training. In the extinction test, it was found that the massed group but not the distributed group showed sensitivity to the changed value of the reinforcer. These results established that instrumental responses that have been reinforced partially can also be sensitive to changes in reward value and that the number of responses emitted per se is necessarily not a determinant of this sensitivity. Furthermore, the findings suggested that the distribution of training may be a mediating factor in the development of habits. Animals received equivalent response-reward pairings but those

that were trained over 10 days failed to show a devaluation effect. In a further study, Adams (1982) examined whether the decline in sensitivity to outcome devaluation over the course of training is the result of greater response emission or exposure to the reinforcer. Animals were trained to press a lever for a sucrose reward. After initially earning 50 reinforcers, one group (low) earned a further 50 in one session on a VR10 schedule, while another group (high) received nine sessions on a VR10 and hence earned a total of 500 rewards. Significantly a third group was yoked to the high group so that after the initial CRF session, it received 9 sessions of non-contingent reward deliveries with the levers retracted and then finally was required to make 500 responses on a VR10 schedule. Thus the yoked group received the same exposure to the reinforcer as the high group, yet experienced the same number of response-reward pairings as the low group. Subsequently, an aversion to the reinforcer was induced in half the animals prior to extinction and reacquisition testing. At test, the low group but not the high group showed sensitivity to outcome value. Critically, the yoked group also failed to show an effect of devaluation. Thus, these results replicated the finding that extended training produces behaviour that is impervious to reinforcer devaluation but suggested that is partly the consequence of the degree of exposure to the reinforcer rather than response repetition per se.

The extent of exposure to the reinforcer is, however, not the only factor that has been shown to contribute to instrumental performance that is independent of the current value of the instrumental outcome. The reinforcement schedule employed during training has also been indicated to differentially affect susceptibility to devaluation treatments

(Adams, 1980; Adams and Dickinson, 1981b; Dickinson, Nicholas and Adams, 1983). Interval and ratio schedules set up distinct causal relationships between an animal's behaviour and reward delivery. On a ratio schedule there is a direct relationship between response and reinforcement rates so that if an animal increases its response rate, there is a concomitant increment in the reinforcement rate. Conversely interval schedules are time dependent and responses are rewarded after a certain amount of time has elapsed since the previous response was made. Dickinson et al. (1983) directly compared the performance of animals trained either on a ratio or interval schedule in a reinforcer devaluation task. After initial training on a CRF schedule, animals received training either on a random ratio (RR) or a random interval (RI) schedule for 2 sessions, earning 30 reinforcers per session. In order to control for difference reinforcement probabilities arranged by these schedules, a third variable ratio group (VR) was yoked to the RI so that the probability of a response being rewarded for the VR group was the same as the RI group. Similarly, to account for any differences in inter-reinforcement intervals (IRI), a fourth VI group was yoked to the RR group and hence the IRI for this VI group would be set by the RR group. Nevertheless, neither of these factors, reward probability or IRI, appeared to contribute to the differential performance of animals in the extinction test. Animals trained on ratio schedules (VR and RR) showed sensitivity to outcome devaluation and pressed the lever at rates significantly lower than non-devalued controls. However, interval schedules appeared to abolish sensitivity to reward value as these animals (VI and RI) showed no devaluation effect. These findings can, therefore, be taken as evidence that training on interval schedules promotes the more rapid onset of behavioural autonomy compared to ratio schedules.

Nevertheless, not all studies of the effects of outcome devaluation on instrumental performance concur with the above findings that extended response practice promotes the onset of behaviour that is habitual and autonomous of the current value of the reinforcer. Indeed, it has been reported that instrumental responding remains sensitive to manipulations in goal-value despite extended training (Colwill and Rescorla, 1985; Colwill and Triola, 2002) and one study has suggested that over-training instrumental responses actually increases sensitivity to outcome devaluation relative to moderately trained animals (Colwill and Rescorla, 1988). These reports of maintained sensitivity to reward devaluation differed markedly in terms of both design and methods from other experiments examining the effect of extended response practice on susceptibility to changes in outcome value (e.g. Adams, 1982). Significantly, these experimenters employed within subject designs with multiple responses and rewards. For example, Colwill and Rescorla (1985) trained animals to make two instrumental responses reinforced with one reward. One response was trained moderately and the second extensively. Moreover, animals were also trained to nose poke for a second reinforcer. Subsequently, gastric malaise was induced in the animals by pairing one of the reinforcers with LiCl and the animals' willingness to respond on both the moderately and extensively trained manipulanda was probed in two extinction tests. These tests revealed that there was no differential impact of the amount of training received on sensitivity to outcome devaluation: both responses were equally sensitive to the devaluation treatment. This use of multiple responses and reinforcers may account for the discrepancy between these results and findings demonstrating behaviour that is autonomous of the current

value of the reinforcer after extended training. Dickinson, Balleine, Watt and Gonzalex. (1995) have argued that under such circumstances the perceived correlation between response emission and reward delivery is maintained and hence behaviour remains sensitive to its consequences. Nevertheless, whatever the explanation of these results, they in no way detract from the large body of evidence that indicate that over-training of a single response-reward pairing can render that response independent of outcome value.

For instrumental behaviour to be characterised as goal-directed it requires not only that the animal has a representation of the reward as a goal but also that instrumental performance be controlled by knowledge of the causal relationship between the action and the outcome (Dickinson and Balleine, 1993). Conversely a behaviour that is impervious to the instrumental contingency between the response and reward delivery can be deemed a habit. As outlined above, sensitivity to the degradation of this contingency by the introduction of omissions schedules is another behavioural assay that can be employed to probe the nature of instrumental conditioning. Dickinson, Squire, Varga and Smith (1998) examined the effect of over-training on sensitivity to omission schedules. Rats were trained to press two levers on a concurrent schedule for pellets. Rats in the low training group earned a total of 120 rewards on each lever and those in the high group, 360. Subsequently, a sucrose solution was presented on a random time (RT) 30-seconds schedule but pressing on the omission lever delayed sucrose delivery. Throughout omission training the contingency between lever pressing and pellet delivery on both levers was maintained. The animals in the low (120) group discriminated between the two levers and suppressed responding on the omission lever. Animals in the

high (360) group, however, failed to discriminate between the two levers and persisted in responding on the omission lever despite the negative contingency. These findings support the position that after extended training behaviour comes to be controlled by S-R habits that are resistant to negative contingencies. Thus over-training of the instrumental response not only engenders behaviour that is insensitive to reward value, it is no longer guided by knowledge of the contingency between the response and outcome delivery.

Taken together, these studies examining the effect of extended response practice on sensitivity to outcome devaluation and degradation of the instrumental contingency provide evidence for a dual-process theory of instrumental performance (e.g. Dickinson 1985; Dickinson et al., 1995). According to this theory, behaviour is initially controlled by A-O associations but over the course of training, S-R processes exert increasing control over behaviour and render it stimulus-bound and habitual. Significantly, since the seminal work conducted by Adams and Dickinson over twenty years ago, this position has received considerable empirical support from studies assessing the impact of outcome devaluation by taste aversion (Yin, Knowlton and Balleine, 2004) and specific satiety (e.g. Killcross and Coutureau, 2003; Coutureau and Killcross, 2003; Faure, Haberland, Conde, and El Massioui, 2005), motivational shifts (Dickinson et al., 1995) as well as contingency degradation (Yin, Knowlton and Balleine, 2006) on the performance of instrumental responses after extended training.

Neuroanatomical dissociation of actions and habits in rats

This behavioural dissociation between response control by goal-directed actions and S-R habits has also been made at the neural level. There is mounting evidence that the basal ganglia, and in particular the dorsal striatum, represent the major neural substrate of S-R and habit learning in rats (Salmon and Butters, 1995; Graybiel, 1995; Packard and Knowlton, 2002).

In part, our understanding of the neural substrates of habit learning has been advanced by dissociating S-R learning in the dorsal striatum from explicit forms of learning mediated by hippocampal and medial temporal cortical systems as well as from stimulus-outcome learning in limbic structures such as the amygdala (White and McDonald, 2002). For example, McDonald and White (1993) showed that animals with the lesions to the dorsal striatum but not the amygdala or hippocampus formation were impaired in their ability to enter an arm of a radial maze for reward when a win-stay strategy was required, but were unimpaired on a standard win-shift task and developed normal conditioned place preference. A similar study in the water maze provided evidence for a dissociation between hippocampal dependent place learning and dorsal striatal S-R learning: animals with dorsal striatal lesions relied on place learning to solve the water maze task but hippocampal lesioned animals were able to associate a cue with a response (McDonald and White, 1994). Indeed, there is evidence that over the course of training animals switch from place to response strategies. This shift to more inflexible and response-

specific performance is suggestive of habitual behaviour and has been shown, critically, to be blocked by inactivation of the dorsal striatum (Packard and McGaugh, 1996).

These studies suggest that S-R learning can be dissociated neuroanatomically from other forms of learning and significantly, is dependent on plasticity in the dorsal striatum. However, it has become increasingly recognised that the dorsal striatum is a highly heterogeneous structure both in terms of its neurochemical compartmentalisation (Graybiel, 1990) and divergent anatomical connections. For example, the main input to the dorsolateral striatum (DLS) arises from the primary sensorimotor cortex but the dorsomedial striatum (DMS) receives projections from the prefrontal and parietal association cortices (Graybiel, 1998; Yin and Knowlton, 2006). There is evidence that this neuroanatomical heterogeneity supports different learning and memory functions (e.g. Reading, Dunnett and Robbins, 1991). Indeed, several studies have dissociated the effects of DLS and DMS on place and response learning in the water maze and report that DLS but not DMS lesions disrupt response strategies in these tasks (Devan and White, 1999; Devan, Mcdonald and White, 1999; Yin and Knowlton, 2004). Similarly, lesions to the DLS have been shown to impair acquisition of an operant conditional discrimination task, while animals with damage to the DMS were unimpaired on this task (Featherstone and McDonald, 2004). These results are consistent with the notion that the DLS is crucial for the acquisition of S-R associations.

The problem with these studies, however, is that the associative structure underlying these apparent forms of S-R learning was not probed. While these behaviours may appear

to be S-R habits, and the authors have assumed them to be so, no empirical data is presented to attest to the veracity of these claims. Furthermore, a rat traversing a maze may operationally be instrumental but it is clear that classical contingencies may contribute to any observed behaviour (Mackintosh, 1983). Nevertheless, understanding of the neural substrates of habit learning has been greatly advanced in recent years through the application of the behavioural assays of instrumental performance outlined in the previous section. These tools allow behaviour that is demonstrably (not just theoretically) habitual to be dissociated from other forms of behaviour and can therefore can be used to analyse the role of distinct neuroanatomical structures in S-R habit learning. Yin et al. 2004 compared the effects of pre-training lesions to the DLS and DMS in an instrumental devaluation task. Animals were trained to press a lever for a sucrose reward on interval schedules and received sufficient training to render performance independent of goal-value. Following instrumental training, an aversion to the sucrose reinforcer was induced in half the animals by LiCl and then the animals' propensity to press the lever was assessed in an extinction test. As expected, sham operated animals pressed the lever irrespective of the devaluation treatment. However animals with lesions to the DLS, but not the DMS, reduced responding after outcome devaluation. The finding of preserved outcome expectancy in animals with lesions to the DLS, in spite of a training regime designed to engender habits, provides the first direct evidence for the involvement of the DLS in habit learning. Furthermore, inactivation of the DLS has been shown to enhance sensitivity to omission schedules compared to controls (Yin et al., 2006). In this study, animals learned to press a lever over 10 days and were then subject to an omission schedule, whereby pressing the lever actually delayed

reward delivery. Prior to the omission training, half the animals received infusions of the GABA-A receptor agonist musimol into the DLS, a procedure that effectively inactivates a neural structure. Consistent with previous findings, extended training produced behaviour that was impervious to the introduction of an omission schedule but inactivation of the DLS appeared to enhance animals' sensitivity to the degradation of the action-outcome contingency. As discussed in the previous section, insensitivity to degradation of the action-outcome contingency is a cardinal feature of S-R habits and as such, this finding highlights the pivotal role of the DLS in habit learning and indicates that this structure is important for maintaining habitual responding during omission learning.

Moreover, these results provide strong empirical support for the existence of two distinct neural systems involved in the control of behaviour by goal-directed actions and S-R habits. Indeed, in stark contrast to the effects of lesions to the DLS, both pre-and post-training lesions to the posterior DMS render instrumental performance habitual, as evidenced by insensitivity to both outcome devaluation and contingency degradation (Yin, Ostlund, Knowlton and Balleine, 2005a; Yin, Knowlton and Balleine, 2005b). An equivalent dissociation has also been found in the medial prefrontal cortex (Killcross and Coutureau, 2003). Killcross and Coutureau compared the performance of animals with either pre-training lesions to the dorsal prelimbic cortex or the ventral infralimbic cortex on a reinforcer devaluation task after limited and extended training. In line with previous reports (Balleine and Dickinson, 1998), animals with lesions to the prelimbic cortex displayed insensitivity to devaluation by specific satiety after limited and extended

training. Conversely after lesions to the infralimbic cortex, instrumental performance remained sensitive to goal value despite extended response practice. After pre-feeding with the instrumental but not an alternative reinforcer, animals with lesions to the infralimbic cortex showed a marked decrement in responding. This disruption of habit formation would suggest that the infralimbic cortex is involved in the process whereby S-R habits come to dominate instrumental performance. The finding that post-training inactivation of this region restores goal-directed performance in overtrained rats provides yet further support for this position (Coutureau and Killcross, 2003). These results suggest that A-O associations are not lost over the course of instrumental training but are actively suppressed by a mechanism that allows S-R habits to achieve behavioural expression. The infralimbic cortex has been implicated in inhibitory control and damage to this region has been associated with a decline in the influence of prior learning on responding (e.g. Quirk, Russo, Barron and Lebron, 2000; Rhodes and Killcross, 2004). This raises the possibility that the role of the infralimbic cortex in the development of habits is to suppress goal-directed actions and modulate competition between the goal-directed and habits systems. There is some, albeit rather weak, evidence of direct anatomical connections between the infralimbic cortex and the dorsolateral striatum (Bayer, 1990; Takagishi and Chiba, 1991) and hence the precise nature of the interaction between these two structures to control habitual responding requires further empirical investigation.

These results underscore the importance of analysing behaviour in terms of the underlying associative processes that support instrumental conditioning. Thus through the

application of modern behavioural assays that unambiguously probe the content of learning, considerable advances have been made in understanding the neural structures involved in the control of instrumental performance after limited and extended training. For example, it has been shown that instrumental responding can be controlled by a goal-directed system involving the prelimbic cortex (Balleine and Dickinson, 1998; Killcross and Coutureau, 2003; Ostlund and Balleine,2005), the dorsomedial striatum (Yin et al., 2005a, 2005b) as well as the nucleus accumbens (Corbit, Muir and Balleine, 2001), mediodorsal thalamus (Corbit, Muir and Balleine, 2003) and basolateral amygdala (Blundell, Hall and Killcross, 2001; Balleine, Killcross and Dickinson, 2003). Similarly, studies have demonstrated that S-R habits are subserved by the dorsolateral striatum (Yin et al., 2004) and the infralimbic cortex (Coutureau and Killcross, 2003). Thus the results from these lesions studies provide support for the existence of two independent and competing neural systems that mediate instrumental performance, and as such, they complement behavioural evidence for a duality of response control by goal-directed actions and S-R habits (e.g. Dickinson and Balleine, 1993). Furthermore, these studies demonstrate that A-O association are not lost as behaviour becomes autonomous of both reward value and the A-O contingency. For example, the finding that post-training inactivation of the dorsolateral striatum (Yin et al., 2006) and the infralimbic cortex (Coutureau and Killcross, 2003) restores goal-directed behaviour in overtrained rats provides clear evidence of intact A-O associations despite over-training. Thus, disruption to one system allows the other system to resume control of behavioural expression. This is significant, as it has been argued that the development of habits reflects a decline in the contribution of A-O processes, and in particular, changes in knowledge about the A-O

contingency (Dickinson, 1985; Dickinson and Balleine, 1993; Dickinson et al., 1995). According to this position, over-training of the instrumental response reduces the experienced correlation between the response and reward delivery and hence degrades knowledge about the A-O contingency (Dickinson, 1985). Similarly as discussed earlier, interval schedules (Dickinson et al., 1983), which establish a non-linear relationship between response and reward rates, and pre-training exposure to the reinforcer (Adams, 1982), have been posited to degrade this relationship and hence promote habit formation (Dickinson, 1985). The findings from lesion studies that behavioural control can switch between two competing systems suggests that the development of habits can not simply be the product of changes in knowledge about the A-O contingency. As yet, however, no behavioural data have been presented to support the notion that disruption to habitual behaviour restores response control to A-O associations. From this perspective, the behavioural literature is at odds with the findings from lesion studies.

Neuroanatomical analysis of habit learning in humans and non-human primates

As with investigations in rats, considerable advances have been made in understanding the neural substrates of habit learning in humans by dissociating it from other forms of learning and memory. Indeed, ever since the demonstration that HM, a patient who received a bilateral resection of the medial temporal lobe resulting in profound amnesia, had intact motor skill learning, the biological basis of multiple memory systems has been recognised (Mishkin, Malamut and Bachevalier, 1984; Milner, Squire and Kandel, 1998). For example, Cohen and Squire (1980) showed that amnesic patients were able to learn a

procedural task of reading mirror-reversed print as well as normal subjects. Such studies have established the existence of dissociable memory systems including declarative explicit memory and non-declarative procedural memory. Declarative memory affords conscious recollection of events and facts, while procedural memory underlies the incremental but unconscious acquisition of skills and the ability to respond automatically to stimuli in the environment.

Basal ganglia dysfunction and in particular damage to the caudate-putamen or neostriatum, the human analogue of the rat dorsal striatum, has been found to be associated with deficits in non-declarative procedural memory (Albin, Young and Penney, 1989; Packard and Knowlton, 2002). For example, patients with Huntington's disease, a genetically transmitted disease producing degeneration of the caudate-putamen, display deficits in tasks, such as the prism adaptation task, that require perceptual-motor skill learning (e.g. Paulsen, Butters, Salmon, Heindel and Swenson, 1993). Significantly, these deficits have been shown to be independent of motor dysfunction in this disease, as Huntington patients are also impaired on probabilistic classification tasks (Knowlton, Squire, Paulsen, Swerdlow and Swenson, 1996a). Probabilistic classification tasks involve learning the relationship between combinations of cues and outcomes and learning the task relies not on explicit memory but rather on a general sense of the rules acquired over numerous trials and as such is potentially a good index of habit learning.

Neuropsychological evidence of the role of the caudate-putamen in human habit learning has been complemented by neuroimaging studies. These experiments have also provided

some evidence of anatomical dissociations between the learning of behaviours that become habitual and the actual production of those responses. For example, two studies of cerebral metabolic activity by positron emission tomography (PET) imaging of subjects learning sequences or executing learned sequences in a button press task have highlighted different patterns of activation during skill learning and the expression of learned responses (Jueptner et al., 1997a; Jueptner, Frith, Brooks, Frackowiak and Passingham, 1997b). These studies revealed that during new learning, there is differential activation of the caudate nucleus and ventral prefrontal cortex but during the performance of already learned sequences there was a posterior shift in activation to the putamen and to the premotor and motor cortex. Moreover, when participants were asked to pay attention to their actions, the caudate and ventral prefrontal cortex were again activated. Similarly, activity in the both the caudate and putamen have been observed while subjects perform a serial reaction time task but the time advantage acquired over repeated task performance is predominately associated with activity in the putamen (Rauch, Whalen, Savage, Curran and Kendrick, 1997). A further study, examining the effects of extended training on a serial reaction time task, revealed decreases in activation in the dorsolateral prefrontal cortex and caudate nucleus, but no equivalent decrease in the putamen, as participants' performance became automatic (Poldrack et al., 2005). Functional magnetic resonance imaging (fMRI) studies of the effect of extended practice of simple motor responses have also revealed a shift with practice from the associative to the sensorimotor striatum (Lehericy et al, 2005). Differential activation of striatal sub-regions over the course of procedural learning has also been observed in non-human primates (Miyachi, Hikosaka, Miyashita, Karadi and Rand, 1997). For example, cell recordings within the

striatum of monkeys performing a sequential button press task have revealed preferential activation of neurons in the caudate nucleus during learning of new sequences but overlearned sequences were associated with neurons in the putamen (Miyachi, Hikosaka and Lu, 2002).

The finding from human imaging studies and neurophysiological studies in non-human primates that as behaviour becomes more automated there is a shift in cortical activation from the caudate nucleus to the putamen is in line with the dissociation found in rats between the medial and lateral dorsal striatum. The DMS and DLS in rats represent the homologues of the human caudate and putamen. Indeed, there is evidence from non-human primate studies of anticipatory, goal-related activity in neurons in the caudate nucleus, suggesting that this structure is involved in goal-expectancy (e.g. Hikosaka, Sakamato and Usui, 1989; Lauwereyns et al., 2002). Furthermore recent evidence from fMRI studies have suggested that the caudate nucleus is not just related to reward processing per se but is involved in the reinforcement of action where there is a contingency between subjects' responses and reward delivery (Tricomi, Delgado and Fiez, 2004; Zink, Pagnoni, Martin-Skurski, Chappelow and Berns, 2004). These studies found caudate nucleus responses to monetary reward, as measured by fMRI, when the receipt of reward was dependent on subjects' performance but not when rewards were delivered non-contingently. These results are consistent with reports that the DMS is a crucial substrate for A-O contingency learning (e.g. Yin et al., 2005a).

Given this anatomical and functional heterogeneity, it is no longer tenable to claim that the dorsal striatum as a whole is a substrate of habit learning. Instead, the more medial caudate nucleus has been shown to be involved in the learning of new sequences of behaviour as well as the modulation of behaviour by outcome expectancy and the contingency between actions and outcomes. Plasticity within the putamen, on the other hand, would appear to be required for the performance of automatic sequences of behaviours that are typical of habits. Indeed, the demonstration of disrupted habit formation after lesions to the dorsolateral striatum in rats (Yin et al., 2004) supports the functional localisation of habits in primates to the putamen. Significantly the transition in neural activity within the primate striatum from the caudate to the putamen, as behaviour becomes more automatic, mirrors the shift in the control of instrumental performance in rats from goal-directed actions to S-R habits over the course of training. These findings suggest that the transition to habitual automatic responding involves dynamic reorganization of neuronal activity within sub-regions of the striatum (e.g. Jog, Kubota, Connolly, Hillegaart and Graybiel,, 1999).

Neurochemical analysis of habit learning in rats

Despite the considerable advances made in recent years in elucidating the neuroanatomical substrates of habit learning, less progress has been achieved in understanding the neurochemical modulation of theses processes. In view of the significance of dopamine in learning and reward (e.g. Waelti, Dickinson and Schultz, 2001; Wise, 2004), it is not surprising that most work examining the neurochemical basis

of habits, has focused on the role of dopamine in S-R habit learning. Significantly, the basal ganglia are characterised neurochemically by prominent input from midbrain dopaminergic pathways. In particular, the dorsal striatum receives dopaminergic innervation form the nigrostriatal dopamine pathway, originating in the substantia nigra pars compacta and contains both D_1-like and D_2-like receptor sub-types (Graybiel, 1990; Bentivoglio and Morelli, 2005).

In parallel to the work dissociating hippocampal and striatal dependent learning reviewed earlier, similar work has found differential effects of dopamine manipulations on these distinct forms of learning. For example, Packard and White (1991) compared the effects of the indirect dopamine agonist amphetamine, the D_1 agonist SKF-38393 and the D_2 agonist LY 171555 infused into the caudate nucleus or the hippocampus on acquisition of win-stay and win-shift tasks in the radial arm maze. Win-stay performance, which is argued to depend on S-R learning, was enhanced by all three dopamine agonists when infused into the caudate-nucleus but not into the hippocampus. Spatial learning on the other hand, as indexed by win-shift behaviour, was improved by hippocampal infusions. Similarly, post-training intracaudate amphetamine infusions have been shown to enhance memory on a cued but not a spatial water maze task (Packard and McGaugh, 1994; Packard and Teather, 1998). Furthermore, intra-striatal infusions of the dopamine antagonist alpha-flupenthixol have been shown to attenuate radial arm maze learning in rats, a task that has be argued to depend on S-R learning (Legault, Smith and Beninger, 2006). These studies suggest that forms of learning and memory that are subserved by the neostriatum are enhanced by dopamine manipulations and as such provide indirect

evidence of dopamine modulation of habit learning. However, none of these studies probed the associative structure underpinning performance on the maze tasks and hence it can only be inferred that the behaviour was truly habitual. Furthermore several of these studies examined the effects of dopamine manipulations in the DMS, a structure which has subsequently been implicated in A-O rather than S-R learning (e.g. Yin et al., 2005a). In view of these considerations, any conclusions from these studies regarding the role of dopamine in S-R habit learning must therefore be tentative.

However, some studies have reported effects of dopamine manipulations on tasks that are not simply assumed to rely on S-R habit learning. Robbins, Giardini, Jones, Reading and Sahakian (1990) examined the effects of systemic administration of the dopamine antagonist flupenthixol and 6-hydroxydopamine (6-OHDA) lesions of the caudate-putamen on the acquisition and performance of a difficult conditional discrimination task. It was found that both systemic and local dopamine depletion in the caudate-putamen disrupted both the acquisition and performance on this task. Significantly, these authors showed that pre-feeding control animals with the instrumental outcome had little impact on well-trained discrimination performance and this suggests that once acquired, this task is not controlled by goal-expectancy and can therefore be characterised as an S-R habit. These results would suggest that dopamine depletion from the dorsal striatum disrupts the neural system responsible for habits and is consistent with a role for dopamine in S-R habit learning. The finding that dopamine depletion affected both the acquisition and performance of the task does not preclude, however, the possibility that other factors such

as motivational, attentional and sensorimotor factors may have contributed to the deficits observed.

A recent study examined the effects of 6-OHDA lesions of the nigrostriatal dopamine pathway on the ability of rats to display behavioural autonomy after extended instrumental training (Faure et al., 2005). As mentioned earlier, the nigrostriatal pathway represents the major source of dopaminergic innervation of the dorsal striatum and destruction to it produces striatal cell loss and dopaminergic deafferentation. Rats learned two different instrumental actions (lever press and chain pull), in the presence of discriminative stimuli, for two distinct rewards. After extended training, the animals' sensitivity to goal value was assessed in a satiety-specific devaluation test so that half the animals were prefed the outcome associated with the lever and the other half outcome paired with the chain. 6-OHDA lesions to the nigrostriatal pathway appeared to disrupt habit formation as both lever pressing and chain pulling was sensitive to the current value of the reinforcer. As such, these results provide evidence for a role in striatal dopamine in habit formation. The problem with this study, however, is that in sham-operated animals only lever pressing appeared to be autonomous of goal-value while chain pulling remained sensitive to the devaluation procedure. This failure to detect habitual responding in half the sham operated animals renders interpretation of the performance of the lesioned animals problematic. Indeed, while it has been consistently shown that over-training of a single action-outcome pairing renders it impervious to outcome value, choice procedures involving multiple responses and rewards can attenuate habit formation (Colwill and Rescorla, 1988; see also Dickinson et al., 1995). It would seem

intuitive that choice procedures are resistant to habit formation because, by their very nature, they require an explicit representation of the outcome to make a choice between two competing actions.

These studies, therefore, provide far from unequivocal evidence for dopaminergic involvement in habit learning. This is not to say that dopamine is not implicated in habit learning, as the lack of evidence is more the product of shortcomings at the level of behavioural analysis in the above-mentioned studies. Evidence from neurophysiological and neurochemical studies demonstrate that striatal dopamine transmission and dopamine-dependent synaptic plasticity is fundamental to learning in the mammalian brain of behaviours that can become habitual (e.g. Reynolds, Hyland and Wickens, 2001; Wickens, Reynolds and Hyland, 2003). The understanding of the role of dopaminergic modulation of S-R learning would be greatly advanced, therefore, by examining the effect of manipulations on instrumental behaviour that be demonstrated empirically to be under the control of the habit system.

Neurochemical analysis of habit learning in humans and non-human primates

Direct evidence for a significant involvement of dopamine in habit learning comes from neuropsychological studies of patient's with Parkinson's disease. Parkinson's disease is a neurodegenerative, hypokinetic movement-disorder but it is also associated with cognitive dysfunction and in particular impaired procedural learning and memory (e.g. Saint-Cyr, Taylor and Lang, 1988). Significantly, Parkinson's disease is characterised by

atrophy of dopaminergic neurons in the substantia nigra and leads to a progressive depletion of striatal dopamine (Albin et al., 1989). For example, Knowlton, Mangels and Squire (1996b) compared the performance of amnesic patients and nondemented patients with Parkinson's disease on a probabilistic classification task which is gradually and implicitly acquired. Participants had to learn to predict which of two outcomes would occur given the particular combination of cues that appeared. Parkinsonian patients, but not amnesics, were severely impaired on the task despite having intact declarative memory for the task. Critically, these effects have been widely replicated and are independent of motor dysfunction in this disease (e.g. Roncacci, Troisi, Carlesimo, Nocentini and Caltagirone, 1999; Hay, Moscovitch and Levine, 2002). The finding that Parkinsonian patients have deficits in tasks that require incremental and implicit learning, characteristic of habits, demonstrates that dopaminergic innervation of the striatum via the nigrostriatal pathway is essential for the development of normal procedural memory.

Further evidence for dopaminergic involvement in human procedural memory also comes from pharmacological studies in healthy participants. Kumari et al. (1997) investigated the effects of the indirect dopamine agonist, amphetamine, and the non-selective dopamine antagonist, haloperidol, in normal volunteers in a procedural learning task that probed implicit and automatic learning. Compared to placebo performance, amphetamine enhanced, while haloperidol attenuated, response speed in this task. Furthermore, the rate of procedural learning observed was related to the degree of dopaminergic manipulation so that the amphetamine-treated subjects displayed the greatest, and subjects administered haloperidol the least, procedural learning. In a further study PET study,

dopamine release was measured while participants played a video game (Koepp et al., 1998). As performance improved on the task, there was decreased binding of radioligand 11C-labelled raclopride (a D_2 receptor antagonist) in the striatum. This decrease in raclopride binding to dopamine receptors in the striatum is consistent with increased release and binding of endogenous dopamine and therefore furnishes yet further empirical support for a role of striatal dopamine in habit learning.

Similarly, the behavioural changes associated with drug abuse are consistent with a role for dopamine in habit learning. Drug-seeking behaviour is initially goal-directed and driven by the reinforcing and hedonic effects of drugs of abuse. Over time, however, there is a loss of voluntary control over behaviour as it becomes progressively more controlled by automatic processes (Tiffany, 1990). It is becoming increasingly recognised that aberrant habit learning may contribute to this process (e.g. Everitt and Robbins, 2005). Furthermore, many of the cognitive deficits associated with chronic drug abuse are characteristic of behaviour that is no longer purposeful and goal-directed. For example, neuropsychological research suggests that chronic drug abusers demonstrate dysfunctional decision-making as indexed by suboptimal performance on gambling and risk assessment tasks. (Rogers et al., 1999; Grant, Contoreggi and London, 2000; Bechara et al., 2001). Similarly, drug addicts have been shown to be impaired on tasks such as the Stroop that require response inhibition (Simon et al., 2000; Salo et al., 2002) and generally have deficits in inhibiting pre-potent responses (Monterosso, Aron, Cordova, Xu and London, 2005). These cognitive deficits are accompanied by behaviour that is no longer controlled by its consequences but rather by its antecedents. For example, drug

cues have been shown to selectively capture attention in human drug addicts (Lubman, Peters, Mogg, Bradley and Deakin, 2000; Franken, Stam, Hendriks and van den Brink, 2003) and exposure to such cues can induce craving and drug-seeking behaviour (Childress et al., 1999; Wang et al., 1999). Significantly, considerable evidence suggests that drugs of abuse influence behaviour as a result of their ability to increase synaptic dopamine, particularly within the striatum (e.g. Wise and Bozarth, 1987; Wise, 1998; Vanderschuren and Kalivas, 2000). As such, the loss of voluntary control over behaviour associated with chronic drug use and the attendant cognitive deficits provide strong evidence for dopaminergic modulation of habit learning.

Moreover, there is good evidence from non-human primate studies that dopamine is involved in the 'stamping in' of S-R habits within the striatum. For example, tonically active neurons (TANs) are progressively recruited as behavioural responses are learnt (Aosaki et al., 1994a) but the acquired responses of TANs are attenuated or eliminated when the nigrostriatal dopamine system is disrupted temporarily by dopamine antagonists (Watanabe and Kimura, 1998) or permanently by unilateral infusions of 1-methyl-4-phenyl- 1, 2,3,6-tetrahydropyridine (MPTP), a dopaminergic neurotoxin, into the caudate-putamen (Aosaki, Graybiel and Kimura, 1994b). These results suggest that dopamine has profound effects on neuronal plasticity within the striatum and as such could affect striatum-based learning and memory. Indeed, it has been suggested that gradual changes in the firing of striatal neurons allows the striatum to 'chunk' representations of action sequences and this process represents a mechanism for the acquisition and performance of action repertoires typical of habits (Graybiel, 1998;

Barnes, Kubota, Hu, Jin and Graybiel, 2005). Disruption to this process by impaired dopamine transmission could contribute to dysfunctional habit learning seen in Parkinson's disease.

Theoretical approaches to the development of habits

Behavioural approaches

Arguably, the most influential theoretical account of the onset of behavioural autonomy over the course of instrumental training has been expounded by Dickinson (Dickinson, 1985; Dickinson and Balleine; 1993; Dickinson et al., 1995). Central to this theory is the idea that animals are sensitive to the experienced correlation between response and reward rates. This was espoused as a critical determinant of instrumental learning by Baum (1973). There is good evidence that animals are sensitive to correlations between reward and response rates. For example, animals will distribute responding between concurrent, but different, interval reinforcement schedules in order to match the relative rates of reinforcement (Herrnstein, 1970). Similarly, the ubiquitous finding that ratio schedules, where there is a linear behaviour-reward function, establish higher rates of responding than interval schedules supports the claim that animals are sensitive to the experienced correlation between behaviour and reward (e.g. Dawson and Dickinson, 1990). When learning an instrumental response animals typically display a negatively accelerating acquisition function so that during the initial stages animals show large changes in response rates across sessions but over time the extent of such changes

diminishes as animals reach some asymptote level of performance. As a result of this negatively accelerating acquisition function, animals may initially experience considerable variation in the correlation between response and reward rates but this experience becomes more restricted over the course of training. According to Dickinson (1985), while animals experience the correlation between behaviour and reward, they encode knowledge about this relation and are therefore sensitive to changes in outcome value. As moment-by-moment experience of this relation declines over the course of training, animals become insensitive to knowledge about the instrumental contingency and hence it no longer influences behaviour and performance becomes habitual.

According to this position, actions become habits as a result of changes in knowledge about the A-O relation. Over-training of an instrumental response degrades this relationship and hence renders performance impervious to goal-value. Similarly, it has been claimed that interval schedules promote the establishment of habits by degrading the continuing experience of the A-O contingency. As is clear from Figure 1, the reward rate as a function of response rate differs markedly across these two classes of reinforcement schedules. The direct relationship between response and reward rates on ratio schedules means that as an animal increases its behavioural output, there is an attendant increase in the reward rate. On interval schedules this relationship is non-linear and very rapidly the reward rate becomes relatively unaffected by anything other than gross changes in the animal's response rate. Thus, on interval schedules animals should typically experience less variation in the correlation between behaviour and reward than animals on ratio schedules. Significantly, training on interval schedules has been shown to engender

habitual responding more rapidly than ratio schedules (Dickinson et al., 1983). Dickinson (1985; Dickinson et al., 1995) has argued that the differential sensitivity to outcome devaluation generated by ratio and interval schedules is attributable to differences in the experienced correlation between response and reward rate on these schedules, and in particular due to greater variation in this factor on ratio schedules.

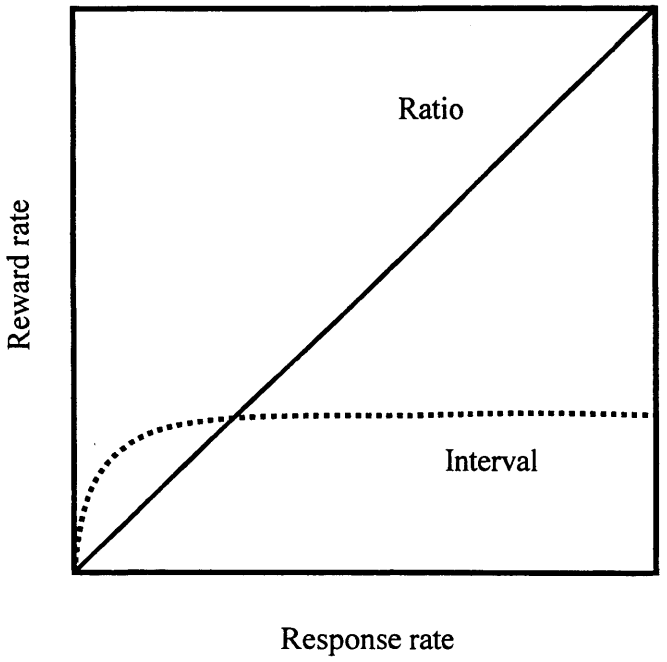

Figure 1. Reward rate as a function of response rate on ratio (filled) and interval (dashed) reinforcement schedules.

This theory is both perspicuous and intuitively appealing and has been very influential, even leading some authors to suggest erroneously that habits cannot be formed on ratio schedules (e.g. Yin and Knowlton, 2006). The theory relies on the assumption that the

experienced correlation, particularly on interval schedules, between behaviour and reward does decline as training proceeds. To date, however, there is no empirical data to support this position. Furthermore, on ratio schedules there is always a strong positive correlation between response and reward rates but training on ratio schedules can engender instrumental performance that is impervious to outcome devaluation procedures. For example, Killcross and Coutureau (2003) trained animals on both interval and ratio schedules but critically the final 5 sessions were on a VR20 and animals showed no sensitivity to outcome value in a subsequent extinction test following devaluation by specific satiety. Moreover, evidence from lesion studies suggests that the process whereby S-R habits come to dominate behavioural expression is the result of a far more active and dynamic process than perhaps previously envisaged. As discussed earlier, the demonstration that disruption of the habit-system restores response control to the goal-directed system suggests that over the course of instrumental training there is a decline in the influence of *intact* A-O associations. If A-O associations are not lost as instrumental training proceeds, it can no longer be held as axiomatic that habits develop as the result of changes in an animal's knowledge of the instrumental contingency.

Neurobiological approaches

A recent theoretical paper has attempted to combine the above behavioural analysis with findings from investigations of the neural substrates of habit learning (Yin and Knowlton, 2006). The authors endeavour to map the behavioural substrates of associative learning onto discrete regions within the brain. Thus in line with results from lesion studies, A-O learning is mapped onto the DMS and S-R habit learning onto the DLS. In order to account for the effects of lesions to the medial prefrontal cortex and the mediodorsal

nucleus of the thalamus on instrumental performance, the authors propose a cortico-basal ganglia network that mediates goal-directed and habitual behaviours. This approach envisages two major networks: the associative and sensorimotor networks. The associative network comprises the mediodorsal thalamus projection to the medial prefrontal cortex which itself projects with the parietal associative cortices to the associative striatum and the DMS. This network is crucial for the acquisition and performance of goal-directed actions but over the course of habit formation control is ceded to the sensorimotor network. The sensorimotor network involves the sensorimotor striatum or DLS, its projections from the sensorimotor cortices and its eventual output targets in the motor cortices. The organisation of the basal ganglia in terms of parallel and re-entrant loops (Alexander, Delong and Strick, 1986) allows interaction between these networks and the authors see this as fundamental to the transformation of behaviour from action to habit. Indeed, the ascending spiral circuitry of the striatum, by which each distinct striatal region influences its adjacent region from the nucleus accumbens shell to the dorsolateral striatum, in a ventral to lateral sequence, provides an anatomical interface for the hierarchical flow of information between these networks (Haber, Fudge and McFarland, 2000). Similarly, work has emphasised the open-interconnected scheme of the basal ganglia-thalamcortical circuitry (Joel and Weiner, 1994; Joel and Weiner, 2000). Such a scheme would permit interactions between networks as well as allowing activity in one network to influence the next and hence facilitate the transfer of behavioural control from the associative to the sensorimotor network.

The proposed hierarchical cortico-basal ganglia network accounts for much of the data. As reviewed earlier, there is strong evidence that habits develop as behaviour comes to be controlled the sensorimotor network described by Yin and Knowlton and originally proposed by Groenewegen et al. (1991). Similarly, there is good evidence that striatal neuronal activation undergoes dynamic reorganisation over the course of habit learning (e.g. Jog et al., 1999; Barnes et al., 2005) and that there is a shift in activity from the associative to sensorimotor striatum (Graybiel, 1998). These data are consistent with the proposal that habit formation leads to the preferential recruitment of the neural circuits outlined in Yin and Knowlton's sensorimotor network. However, the authors fail to incorporate into the model evidence that the infralimbic cortex forms an integral part of the neural substrates of habit learning (e.g. Killcross and Coutureau, 2003). Furthermore, one of the fundamental questions lies with how the transition from goal-directed actions to habits actually takes place and the extent to which the systems interact or compete to achieve behavioural control. To answer this question the authors evoke Dickinson's account of this transition to habits in terms of experienced contingency. It is suggested that there are neural detectors for response and outcome rates and that these detectors calculate an estimate of experienced contingency. Over-training reduces the output of the contingency detector and as a result the A-O system relinquishes control to the habit system. The neural substrates and mechanism of action of this supposed contingency detector, of course, remain to be elucidated. As such the proposed network merely amalgamates findings from investigations into the neural substrates of habit learning with the behavioural model outlined earlier and provides few, if any, new insights into the development of habits.

Neurocomputational approaches.

Of much more importance is recent theoretical work addressing the key issue of how competition between the goal-directed and habits systems is resolved from a normative perspective by the application of the computational theory of reinforcement learning (Daw, Niv and Dayan, 2005; Daw, Niv and Dayan 2006). Reinforcement learning theory examines adaptation and optimization in the selection of actions (Sutton and Barto, 1998). It is assumed that the goal of all action is to maximize or optimize some measure of utility (e.g. future gain). In reinforcement learning, actions are assessed through predictions of their values, where values represent the ultimate reward (or punishment) associated with any given action. The animal or agent is presumed to use these value functions to predict the rewards that may be yielded by any given action. In order that action selection optimizes long-term utility, value prediction must be based on estimates of the long-term value of an action. There are two major classes of reinforcement learning 'model-free' and 'model-based', both of which utilize different approximations of value. Daw and colleagues argue that these two classes of reinforcement learning subserve goal-directed actions and habits respectively and propose a mechanism to mediate competition between the two systems.

Temporal difference learning is a model-free approach based on evidence that the phasic activity of dopaminergic neurons in the striatum represent errors in the prediction of future reward (e.g. Ljungberg, Apicella and Schultz, 1992; Schultz, 1998). The concept of 'caching', the association of an action with a scalar summary of its long-term future

value, is central to this class of reinforcement learning. Such a learning system only stores anticipated sum value and is independent of the *actual* consequences of a given action. Significantly, it has been shown that over the course of training the firing-response of dopamine neurons is transferred from the reward to reward-predicting stimuli (Schultz, Dayan and Montague, 1997). Cached values are computationally simple but inflexible as values do not change if the outcome is devalued. The inflexibility and insensitivity to changes in reward of learning by caching also represent the cardinal behavioural characteristics of the dorsolateral striatal habit system. The goal-directed system, on the other hand, is subserved by a 'model based' reinforcement learning method. This method involves 'tree search' whereby predictions of value are obtained by exploring the immediate consequences of each action and constructing a model of the long-term consequences of any action by searching the entire model recursively. Tree search is cognitively demanding but as value predictions are based on actual terminal consequences, it is behaviourally flexible and can respond efficiently to changes in circumstances. As such, the model captures the hallmark of goal-directed behaviour, namely sensitivity to outcome revaluation procedures.

These two classes of reinforcement learning theories can produce different estimates of the value of a given action and in order to achieve optimal control over behaviour and hence maximize the probability of reward, arbitration between the two systems is required. It is assumed that arbitration is based on the relative *reliabilities* of the systems and the system associated with the least uncertainty determines action selection. Early on, model search is more efficient at gaining value predictions from small amounts of

experience, but over time there is an increase in uncertainty as extensive tree search adds noise and forces approximation; this renders the cache system less uncertain and hence more advantageous. The authors provide data from simulations showing that over the course of acquisition there is indeed a transition from model-driven to model-free control, which mirrors the shift from R-O to S-R processes observed in animals (Daw et al., 2005). Furthermore, this would account for the failure of training with multiple responses and outcomes (e.g. Colwill and Rescorla, 1985) to promote habitual responding because an animal's experience is spread between more actions and values and tree search is less uncertain than caching when there is such paucity of data regarding action value. Furthermore, one of the considerable advantages of the model is that by proposing two competing and interacting controllers, it is consistent with the results of lesion studies that demonstrate a dissociation between neural systems controlling goal-directed actions and habitual responding. However, the extent to which the goal-directed and habit systems compete or function independently to achieve behavioural control is as yet not clear from existing behavioural and neural data and therefore the suggestion of uncertainty-arbitration between two competing controllers requires empirical verification.

Behavioural and neurochemical analysis of the development of S-R habits in rats: overview of Experiments 1-13.

The experiments conducted for this thesis seek to provide further elucidation of both the behavioural and neurochemical processes involved in the development of S-R habits in rats. As previous research has established that instrumental actions, such as lever

pressing, can be controlled by their consequences (goal-directed actions) or by antecedent stimuli (S-R habits), it is critical to analyse behaviour in terms of the underlying psychological processes that support performance. Thus, central to all the experiments reported here is the application of behavioural assays (reinforcer devaluation) that probe the associative structure underpinning instrumental performance. Sensitivity to outcome devaluation indexes the extent to which outwardly similar behaviours are controlled by goal-directed A-O associations or S-R habits. Experiments 1 and 2 aimed to replicate the ubiquitous finding that overtraining an instrumental action leads to goal-insensitive habitual responding. Significantly, these experiments explored previous suggestions that the class of reinforcement schedule employed during training (interval or ratio) can influence the sensitivity of an instrumental response to changes in outcome value. Furthermore, by examining both local and global response and reward rates, these experiments sought to identify factors critical to the development of S-R habits. These factors were further explored in Chapter 3. In Experiments 3-7, the impact of various manipulations (changes in the reinforcement schedule, reward magnitude, reward exposure) on the sensitivity of an overtrained instrumental was examined. The experiments reported in Chapters 2 and 3 sought not only to highlight the psychological processes involved in the transition from action to habit but also to provide behavioural evidence that even after extended training, response control can revert from the habit to the goal-directed system.

While considerable advances have been made in recent years in elucidating the neuroanatomical circuitry that underpins the performance of purposeful goal-directed

actions and stimulus-bound habits, little progress has been made in understanding the neurochemical processes involved. The experiments reported in Chapters 4-5 (Experiments 8-13) attempt to address this discrepancy. In view of the pivotal role of dopamine in both learning and reward as well as evidence of procedural learning deficits in Parkinson's disease, there are good grounds for exploring dopaminergic modulating of habit formation. Chapter 4 examined the effect of both pre- and post-training exposure to the indirect dopamine agonist amphetamine on the sensitivity of a moderately trained instrumental response to outcome devaluation (Nelson and Killcross, 2006). These experiments suggested that pre-training exposure to amphetamine accelerates the normal incremental process whereby S-R habits come to dominate instrumental performance. The final experiments (Experiments 10-12) in this thesis examined the pharmacological specificity of the sensitization effect observed in Chapter 4. This was achieved through the systemic administration of both selective and non-selective dopamine antagonists during the acquisition of an instrumental response following amphetamine sensitization. Finally in Experiment 13, the different contribution of D_1 and D_2 receptor subtypes to instrumental learning was examined in non-sensitized animals.

CHAPTER 2

Experiments 1-2: The effect of amount of training and reinforcement schedule on the sensitivity of an instrumental response to outcome devaluation by specific satiety.

As reviewed in Chapter 1, a body of evidence has accumulated demonstrating that in rats, instrumental actions such as lever pressing, can be controlled by two dissociable associative structures. Initially performance is guided by voluntary goal-directed actions and animals are sensitive to changes in outcome value. However, over the course of training response control is relinquished to reflexive S-R habits and instrumental performance becomes autonomous of goal-value. To date, the number of reinforcers earned, as well as the reinforcement schedule employed, have been identified as the critical determinants of sensitivity to changes in outcome value (Dickinson et al., 1983). Thus as exposure to the reinforcer increases, there is an attendant decrease in animals' susceptibility to devaluation procedures. Similarly, interval schedules have been posited to promote the more rapid onset of behavioural autonomy compared to ratio schedules. As outlined in the last section of Chapter 1, this process has been argued to reflect a decline in the perceived correlation between response and reward rates (e.g. Dickinson, 1985).

As noted earlier, however, little or no empirical evidence has been presented to support or test this theoretical position on the transition from actions to habits. Thus the aim of the experiments reported here was firstly to provide a replication of the finding that overtraining promotes the development of S-R habits. In view of the evidence that the degree of exposure to the reinforcer and not the number of responses *per se* predicts sensitivity to outcome devaluation, animals earned 120, 240 or 360 rewards either on an interval (Experiment 1) or a ratio (Experiment 2) schedule of reinforcement prior, to a test of sensitivity to outcome devaluation by specific satiety. In an attempt to match reward and response rates across the two experiments and reward schedules, we used the mean lever press rate and reward rate generated on the interval schedule in Experiment 1 to set the mean number of responses required to earn a reward on the ratio schedule in Experiment 2. By varying the amounts of training and reinforcement schedules, these experiments would test the hypothesis that interval and ratio schedules promote differential sensitivity to outcome devaluation. Furthermore, analysis of both local and global response and reward rates during training and subsequent sensitivity to outcome value in the devaluation test would probe the suggestion that a decline in the perceived correlation of these two factors is responsible for the development of behavioural autonomy.

Method and Materials

Subjects

In Experiment 1, the subjects were 48 experimentally naïve male Lister-hooded rats of mean weight 276g at the beginning of the experiment (range 238 – 289g). In Experiment 2, a further 48 experimentally naïve male Lister-hooded rats were used, with a mean weight of 298g (range 247 – 310g). The rats were handled daily and accustomed to the temperature and humidity-controlled laboratory vivarium for 1 week before the start of the experiments. They were housed two per cage. The vivarium was maintained at 21° C with the light on from 8 am to 8 pm. All experiments were conducted in the light phase. Animals were food deprived for one week before the commencement of behavioural training and reduced to 80% of their free-feeding weight. Water was available *ad libitum* throughout. All experimental procedures involving animals and their care were carried out in accordance with the UK Animals (Scientific Procedures) Act 1986 and were subject to Home Office approval (Project License PPL 30/2140).

Behavioural apparatus

The training apparatus comprised 8 chambers (Paul Fray Ltd, Cambridge, UK) measuring 25 x 25 x 22 cm. The chambers were individually housed within sound-attenuating cabinets and were ventilated by low noise fans. Each chamber had three aluminium walls and a clear Perspex front wall. The roof was made of clear Perspex and the floor consisted of 18, 5 mm diameter steel bars spaced 1.5 cm apart centre-to-centre, parallel to the back of the chamber. A recessed magazine that provided access to rewards via a

hinged Plexiglas panel was located in the centre of the left-hand wall. Two reinforcer types could be delivered into the magazine in each chamber. 0.1ml of sucrose solution (15% w.v) via a persilistic pump and food pellets (Formulae A/I, 45mg, P J Noyes, Lancaster, NH) by a pellet dispenser. Levers could be inserted to the left and the right of the magazine. A houselight mounted on the roof provided illumination. The apparatus and on-line data collection were controlled by means of an IBM-compatible microcomputer equipped with MED-PC software (Med Associates Inc. VT)

Behavioural Training

Training consisted of two stages – magazine training and lever pressing. Throughout the training phase both magazine entry and lever press behaviours were measured. Half of the animals were trained with food pellets and the other half with 15% w/v sucrose solution. Training was followed by devaluation by specific satiety and extinction testing.

Magazine training. Animals were first trained to collect rewards during two 30-minute magazine training sessions. The rewards were delivered on a random time 60-second schedule. The levers were not available in these sessions. The duration of each 30-minute session was signalled by the illumination of the houselight. In the first session of magazine training, the Plexiglas panels were pinned back.

Lever press training. Thereafter the levers were inserted into the boxes and all the rats were given 2 sessions of lever press training on a continuous schedule of reinforcement schedule (CRF). In each of these sessions animals earned a total of 25 reinforcers. In

Experiment 1, the animals then received training on a random interval 30 second schedule (RI30). In each of these sessions animals earned a total of 40 rewards. The Low group (n=16) received three of these sessions and thus earned a total of 120 reinforcers on the RI30 schedule. The Medium group (n=16) had 6 sessions on the RI30 schedule earning 240 reinforcers and the High group (n=16) received 9 sessions thus earning 360 reinforcers on this schedule. In Experiment 2 animals were trained on a random ratio 8 schedule (RR8). This schedule was selected because the mean number of responses required in Experiment 1 to earn one reward was approximately 8 (mean response rate per minute 13.75, mean reward rate per minute 1.77). As in Experiment 1, animals earned in total 120 (Low group), 240 (Medium group) or 360 (High group) rewards in total.

Devaluation by specific satiety

All animals then received 1 session of specific-satiety devaluation extinction testing during which lever press and magazine entry behaviours were assessed. On the day after the final day of instrumental training each rat was placed in a feeding-cage and given free access to one of the two types of reward for 60 minutes (half received food pellets and the other half sucrose solution). Half of the animals in each group were prefed the instrumental outcome (devalued group) and the other half the alternative reinforcer (non-devalued). After this pre-feeding session the animals were transferred to the conditioning chambers and received a 10-minute extinction test during which responding was measured in the absence of reward delivery.

Data analysis

Statistical analysis was performed using analysis of variance (ANOVA) with between subject factors of training (with 3 levels high, medium or low) and devaluation by specific satiety (devalued versus non-devalued). Significant main effects with more than two levels are explored with Tukey pairwise comparisons. All statistcal analyses use an alpha level of 0.05.

Results

Experiment 1. The sensitivity of an instrumental response to outcome devaluation by specific satiety after different amounts of training on an interval schedule.

Instrumental training. Figure 2.1.1 displays the mean lever presses per minute on the final day of R130 training. (Day 3 Low group, day 6 Medium group and day 9 High group). It is clear from this figure that all animals, irrespective of devaluation group, acquired the instrumental response at equivalent rates but that responding in the Medium group was marginally elevated compared to the Low group. This observation was supported statistically by ANOVA which yielded a main effect of training ($F_{(2,42)} = 3.86$, $p<0.05$). Subsequent pairwise analysis with Tukey tests confirmed that responding was higher in the Medium group compared to the Low group ($p<0.05$) but no other comparisons were statistically significant. However, there was no effect of intended devaluation nor an interaction between this factor and training (both Fs<1). Analysis of magazine entry data on the final day of instrumental training revealed no effects of

training, intended devaluation or an interaction (highest $F_{(1,42)} = 1.67$, p=0.203) (mean magazine entries per minute (±SD): devalued Low group = 10.2. (± 46.2); Low non-devalued group = 6.1 (± 2.3) ; devalued Medium group = 7.3 (± 2.8); non-devalued Medium group = 6.9 (±6.0); devalued High group = 7.5 (± 2.9); non-devalued High group = 7.3 (± 3.9).

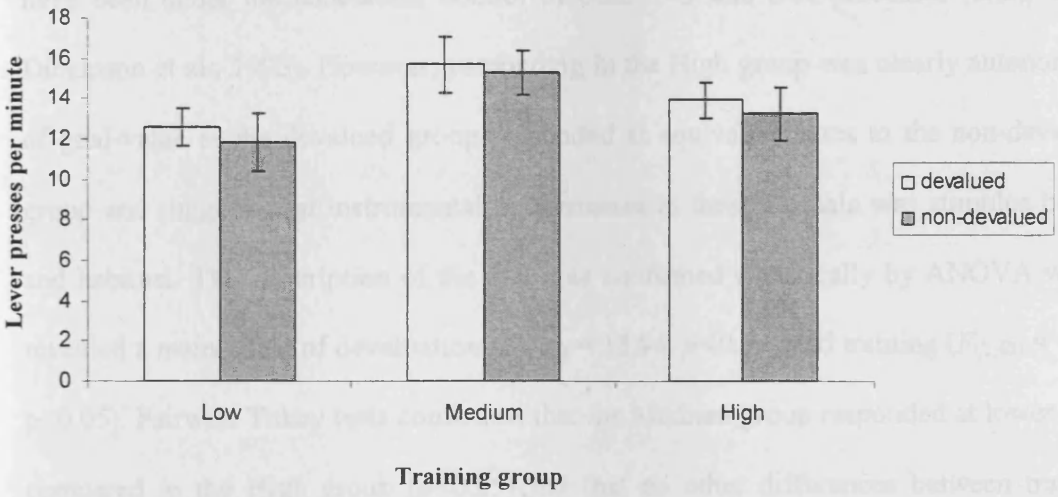

Figure 2.1.1 Mean lever presses per minute (±SEM) on the final day of acquisition after varying amounts of training.

Extinction test – lever press performance. In order to control for baseline differences and reduce within subject variability, the extinction data were analysed as a proportion of baseline performance. These are presented in Figure 2.1.2. The suggestion from this figure is that as the amount of training increased there was a concomitant decrease in animals' sensitivity to outcome devaluation. Responding in the Low group appeared to be goal-directed as evidenced by the marked suppression in lever pressing by animals

prefed the instrumental outcome (white bars) compared to the non-devalued controls (grey bars). Similarly, the Medium group appeared to be sensitive to goal-value albeit to a lesser extent than the Low group. There was also lower responding in the non-devalued Medium group relative to the non-devalued Low group. As motivational manipulations, such as pre-feeding to satiety, are known to exert a non-selective impact on S-R associations, this can be taken as evidence that responding in the Medium group might have been under the concurrent control of both A-O and S-R processes (Hull, 1943; Dickinson et al., 1995). However, responding in the High group was clearly autonomous of goal-value as the devalued group responded at equivalent rates to the non-devalued group and suggests that instrumental performance in these animals was stimulus-bound and habitual. This description of the data was confirmed statistically by ANOVA which revealed a main effect of devaluation ($F_{(1,42)} = 18.94$, $p<0.05$) and training ($F_{(2,42)} = 4.04$, $p<0.05$). Pairwise Tukey tests confirmed that the Medium group responded at lower rates compared to the High group ($p<0.05$) but that no other differences between training groups were statistically significant. Critically, there was also a significant training x devaluation interaction ($F_{(2,42)} = 4.33$, $p<0.05$). Simple effects analysis of this interaction confirmed that responding in the Low group was goal-directed as there was a highly significant effect of devaluation in these animals ($F_{(1,14)} = 21.24$, $p<0.05$). There was also an effect of devaluation in the Medium group ($F_{(1,14)} = 6.6$, $p<0.05$) but none in the High group ($F<1$).

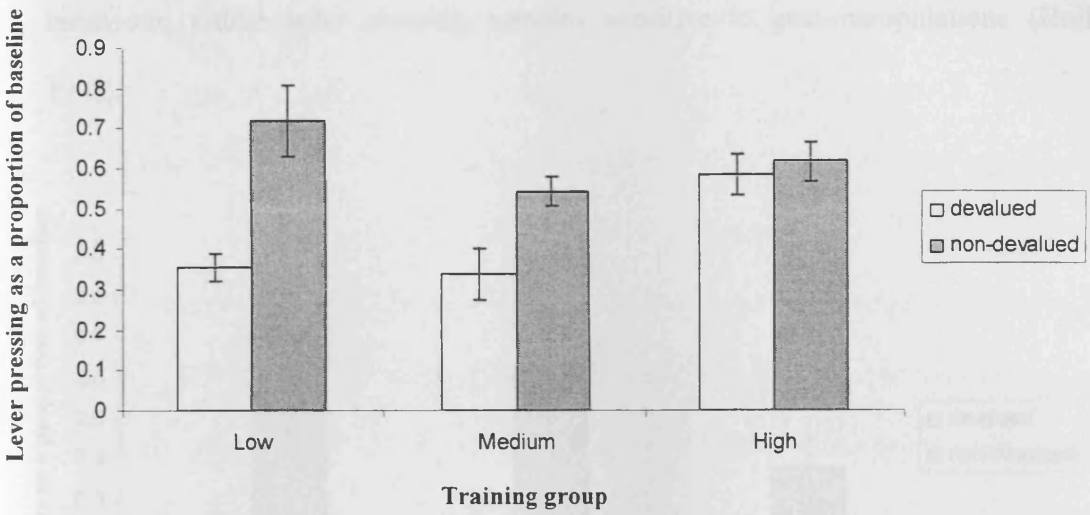

Figure 2.1.2 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after pre-feeding with the instrumental outcome (devalued - white bars) or an alternative reinforcer (non-devalued - grey bars).

Extinction test – magazine entry behaviour. Analysis of magazine entry behaviour during the extinction yielded a main effect of devaluation ($F_{(1,42)} = 14.6$, $p<0.05$). As is clear from Figure 2.1.3, all animals, irrespective of the extent of training, performed fewer magazine entries when prefed the instrumental outcome compared to animals prefed the alternative reinforcer. There was no effect of training ($F_{(2,24)} = 1.61$, $p=0.21$) nor an interaction ($F<1$). Thus, in marked contrast to lever pressing, magazine approach behaviour remained sensitive to outcome devaluation even after extensive training. This finding indicates that the differential sensitivity of lever pressing to outcome devaluation reported above can not be explained by any ineffectiveness of the pre-feeding procedure and is consistent with previous reports that after overtraining magazine approach

behaviour, unlike lever pressing, remains sensitive to goal-manipulations (Holland, 1998).

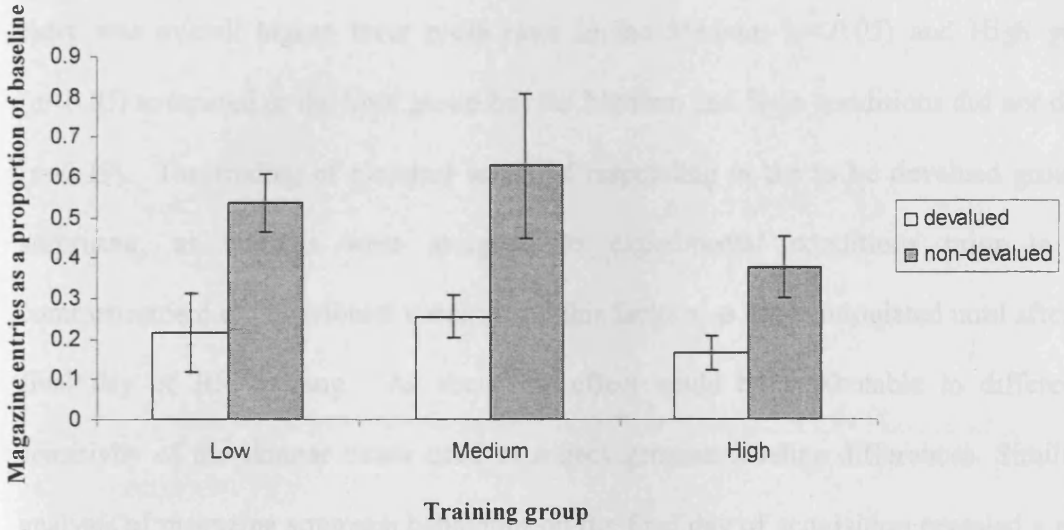

Figure 2.1.3 Mean magazine entries per minute (±SEM) in the extinction test after pre-feeding with the instrumental outcome (devalued - white bars) or an alternative reinforcer (non-devalued - grey bars).

Experiment 2. The sensitivity of an instrumental response to outcome devaluation by specific satiety after different amounts of training on a ratio schedule.

Instrumental training. The mean lever presses per minute on the final day of acquisition for the Low, Medium and High training conditions are presented in Figure 2.2.1. Inspection of this figure indicates that extended training promoted higher response rates. Furthermore, the suggestion from this figure is that there were marginally higher rates of responding in the devalued group compared to the non-devalued group. This description

of the data was supported by statistical analysis. ANOVA yielded a main effect of training ($F_{(2,24)}$ = 23.07, $p<0.05$), intended devaluation ($F_{(1,42)}$ = 6.789, $p<0.05$) but critically no interaction between these factors ($F<1$). Pairwise Tukey tests confirmed that there was overall higher lever press rates in the Medium ($p<0.05$) and High group ($p<0.05$) compared to the Low group but the Medium and High conditions did not differ ($p=0.19$). The finding of elevated levels of responding in the to be devalued group is surprising, as animals were assigned to experimental conditions prior to the commencement of behavioural training and this factor was not manipulated until after the final day of RR training. As such this effect could be attributable to differential sensitivity of the skinner boxes used or reflect genuine baseline differences. Similarly, analysis of magazine approach behaviour on the final day of acquisition revealed a main effect of intended devaluation ($F_{(1,42)}$ = 4.2, $p<0.05$) but no effect of training ($F_{(2,24)}$ = 1.28, $p=0.29$) or interaction ($F<1$) (Mean magazine entries per minute (±SD): Low devalued group = 8.5 (± 4.7); Low non-devalued group = 4.5 (±3.1); Medium devalued group = 9.2(± 2.8); Medium non-devalued group = 7.4 (± 5.2); High devalued group = 8.7 (± 8.7); High non-devalued group = 8.0 (± 3.2)

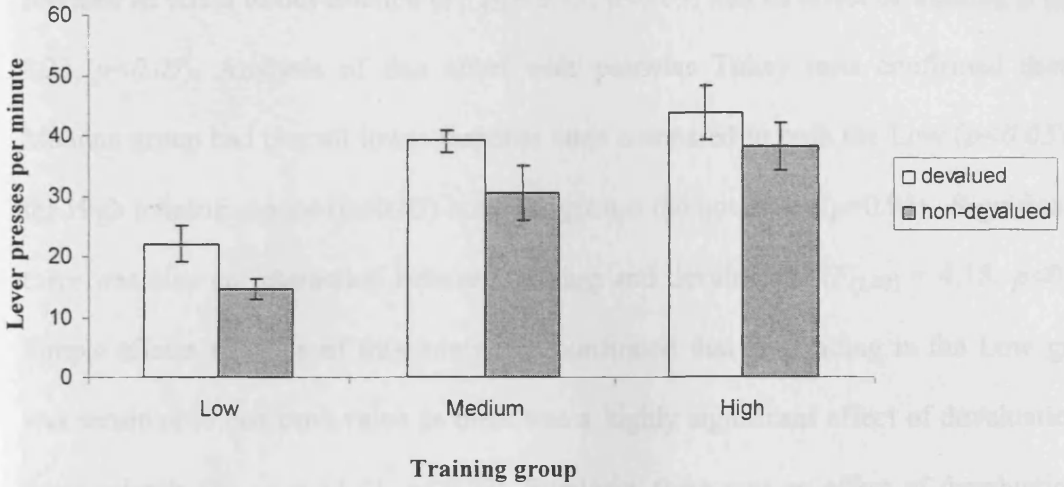

Figure 2.2.1 Mean lever presses per minute (±SEM) on the final day of acquisition after varying amounts of training.

Extinction test – lever press performance. In view of baseline differences in responding, the extinction data were analysed as a proportion of baseline rates and are displayed in Figure 2.2.2. Inspection of this figure indicates that responding in the Low training group was goal-directed as the animals prefed the instrumental outcome performed fewer lever presses per minute as a proportion of baseline compared to the non-devalued controls. Similarly, performance in the Medium training group appeared to be guided by outcome value but there was also relatively low rates of responding in the non-devalued controls and as such can be taken as evidence of a general suppression in responding after pre-feeding to satiety. Conversely, lever press performance in the High training group appeared to be controlled by S-R habits as there was no evidence of an effect of the devaluation treatment in these animals; if anything responding was higher in the animals prefed the instrumental outcome. These observations were supported by ANOVA which

revealed an effect of devaluation ($F_{(1,42)} = 9.73$, $p<0.05$) and an effect of training ($F_{(2,42)} = 8.33$, $p<0.05$). Analysis of this effect with pairwise Tukey tests confirmed that the Medium group had overall lower response rates compared to both the Low ($p<0.05$) and the High training groups ($p<0.05$) but these groups did not differ ($p=0.95$). Significantly, there was also an interaction between training and devaluation ($F_{(2,42)} = 4.18$, $p<0.05$). Simple effects analysis of this interaction confirmed that responding in the Low group was sensitive to outcome value as there was a highly significant effect of devaluation in these animals ($F_{(1,14)} = 11.41$, $p<0.05$). Similarly, there was an effect of devaluation in the Medium group ($F_{(1,14)} = 6.53$, $p<0.05$) but responding in the overtrained animals was impervious to reward value ($F<1$). With respect to the extinction performance of the devalued groups, it is perhaps important to address the issue of the higher baseline rates in these animals compared to the non-devalued controls. As the extinction data were analysed as a proportion of baseline rates, it could be argued that the devaluation effects seen at test are a consequence of these elevated baseline rates. This is, however, most unlikely. The devalued animals in the High group also displayed higher baseline rates compared to non-devalued controls but at test there was no selective depression in lever press performance in animals prefed the instrumental outcome.

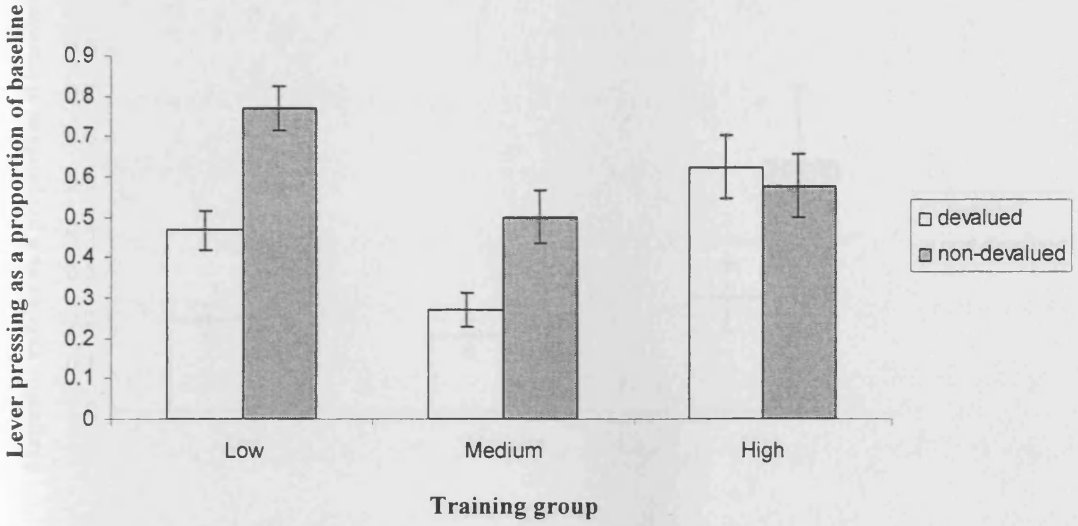

Figure 2.2.2 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after pre-feeding with the instrumental outcome (devalued - white bars) or an alternative reinforcer (non-devalued - grey bars).

Extinction test – magazine entry behaviour. Figure 2.2.3 displays the mean magazine entries per minute as a proportion of baseline during the extinction test. It is clear from this figure that in contrast to lever press performance, magazine approach behaviour remained sensitive to outcome value despite extended training. This was confirmed by ANOVA which yielded a main effect of devaluation ($F_{(1,42)} = 8.46$, $p<0.05$) but no effect of training ($F_{(2,24)} = 1.66$, $p=0.20$) and significantly no interaction between these factors ($F<1$). These data confirm that pre-feeding to satiety had successfully devalued the instrumental outcome for all the training groups.

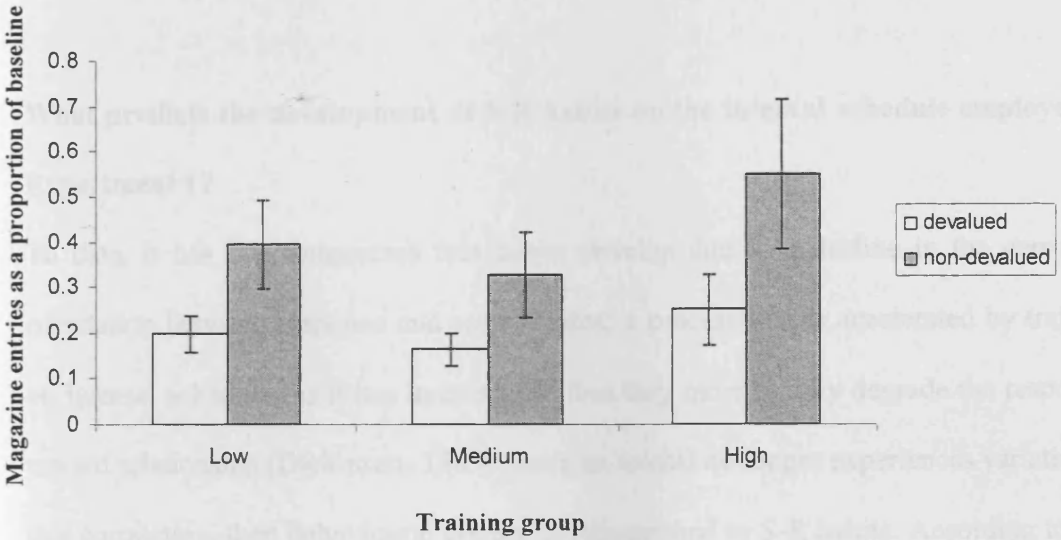

Figure 2.2.3 Mean magazine entries per minute (±SEM) in the extinction test after pre-feeding with the instrumental outcome (devalued - white bars) or an alternative reinforcer (non-devalued - grey bars).

Summary

Experiment 1 and 2 explored the effect of varying amounts of training on sensitivity of lever press performance to changes in reward value after pre-feeding to satiety on the instrumental outcome. Significantly, the experiments replicated the ubiquitous finding that over the course of training, instrumental performance becomes independent of the current value of the reinforcer as response control is ceded from A-O associations to S-R habits. Furthermore in the current experiments, this effect was seen regardless of the reinforcement schedule employed. Here interval schedules did not promote the more rapid onset of behavioural autonomy and nor did ratio schedules attenuate the onset of response control by S-R habits: on both schedules, robust habits developed after equivalent amounts of training.

What predicts the development of S-R habits on the interval schedule employed in Experiment 1?

To date, it has been suggested that habits develop due to a decline in the perceived correlation between response and reward rates; a process that is accelerated by training on interval schedules as it has been argued that they more rapidly degrade the response-reward relationship (Dickinson, 1985). Once an animal no longer experiences variation in this correlation, then behavioural control is relinquished to S-R habits. According to this position, an animal's performance on the response-reward rate curve should predict its sensitivity to outcome devaluation. After limited training animals should still be moving up and down the response-rate curve and hence experiencing the response-reward correlation. Conversely, overtrained animals should no longer be exploring the relationship between responding and reward delivery and hence behaviour should no longer be controlled by knowledge of this relationship. If overtraining of an instrumental response reduces the experienced correlation between behaviour and reward, there should be a relationship between the amount of training received and position on this curve. For example, in the current experiment animals in the High group would be expected to have reached asymptotic performance in terms of the maximum number of rewards available per minute (2). On the other hand, animals that were sensitive to goal-value at test (Low and Medium groups) should have still experienced the response-reward correlation in the final session of acquisition and would not be expected to have earned the maximum number of rewards available per minute on the RI30 schedule. To test this hypothesis,

the position on the response-reward rate curve on the final day of acquisition of the Low, Medium and High groups in Experiment 1 is plotted on Figure 2.3.1.

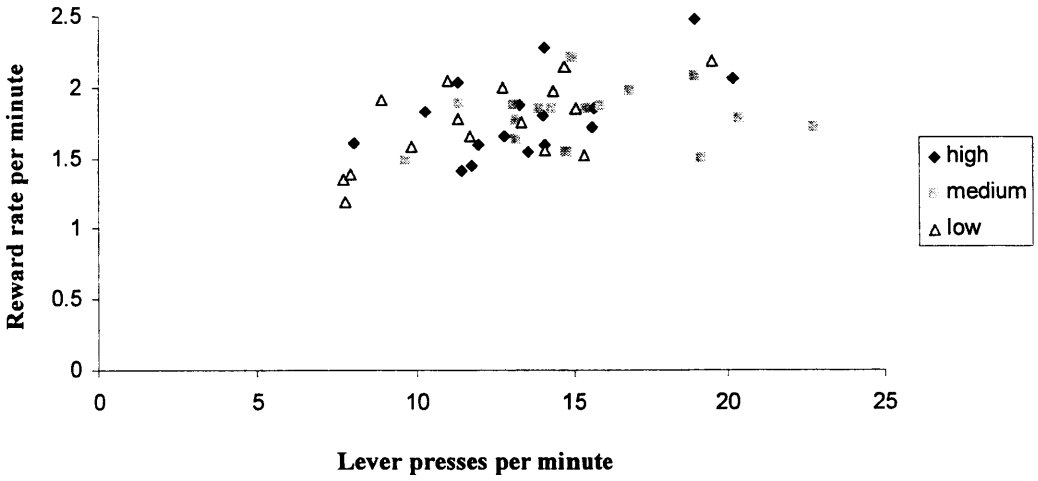

Figure 2.3.1 Reward-response rate function of Low, Medium and High groups on the final day of RI training in Experiment 1.

As is clear from this figure, there was no systematic relationship between position on this graph and the amount of training received. The suggestion from this figure is that there was a trend towards higher reward rates as response rates rose. This was confirmed statistically ($r_{(46)} = 0.46$, $p<0.05$) but it is evident that this trend was irrespective of the amount of training received. Indeed, partial correlations controlling for training group revealed an equally robust relationship between lever press and response rates ($r_{(45)} = 0.45$, $p<0.05$). Thus, animals in the High group had not necessarily reached asymptotic performance in terms of earning the maximum available rewards per minute and could theoretically, therefore, still experience the correlation between lever pressing and reward delivery. Similarly, there was no consistency in the Low and Medium groups in terms of

position on the response-reward curve, with some animals in these groups obtaining the maximum number of rewards available per minute. Thus despite marked sensitivity to outcome value at test, animals in the Low and Medium groups may not have experienced changes in the behaviour-reward correlation as indexed by their position on this curve. As Figure 2.3.1 demonstrates that the behaviour-reward function does not vary according to the amount of training received, it suggests that overtraining of an instrumental response does not necessarily degrade this relationship. As such, these data provide little supprt for the proposal that changes in the behaviour-reward function are critical to the development of S-R habits.

Indeed, further analysis of the relationship between lever press and reward rates in the final training session and sensitivity to outcome devaluation in the subsequent extinction test confirmed that there was no predictive relationship between any of these factors. There was a non-significant trend towards a negative correlation between extinction performance and lever press rates on the final day of acquisition ($r_{(22)}$ = -0.24, p=0.25) but some relationship between these factors would be expected as the extinction data are expressed as a proportion of these lever press rates. Figure 2.3.2 displays the reward rates on the final day of acquisition and subsequent extinction performance for animals in the devalued groups. It is evident from this figure that there is no relationship between the average reward rate in the final session of instrumental training and sensitivity to outcome value after pre-feeding with the instrumental outcome ($r_{(22)}$ = -0.12, p=0.57).

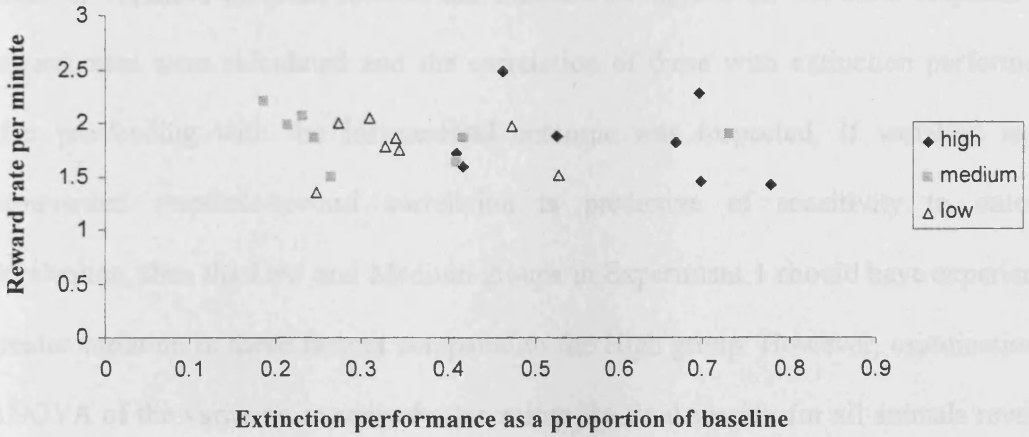

Figure 2.3.2 Reward rates on the final day of acquisition and extinction performance as a proportion of baseline after pre-feeding with the instrumental outcome.

From these data it would appear that it is not tenable to claim that the experienced correlation between lever and response rates predicts subsequent sensitivity to outcome devaluation. There is no evidence from Experiment 1 that animals in the High devalued group experienced less correlation between these factors compared to the Low and Medium groups but these groups still demonstrated differential sensitivity to changes in outcome value. Of course, these correlations are based on the global rates for the final session of RI training. It remains possible that animals in the Low and Medium groups experienced greater variation in the response-reward correlation across the final session and hence this may account for the differential sensitivity to devaluation by specific satiety observed at test.

In order to account for this possibility, local response and reward rates (assessed on a minute by minute basis) during the final session of RI training were recorded. As an

index of variation in these factors, the standard deviations for the local response and reward rates were calculated and the correlation of these with extinction performance after pre-feeding with the instrumental outcome was inspected. If variation in the experienced response-reward correlation is predictive of sensitivity to outcome devaluation, then the Low and Medium groups in Experiment 1 should have experienced greater variation in these factors compared to the High group. However, examination by ANOVA of the variation in reward rates across the final session for all animals revealed that there were no differences between the three training groups ($F<1$) (mean variation in reward rate Low group = 1.12; Medium group = 1.07; High group = 1.15). Significantly, there was no correlation between variation in reward rates and extinction performance in animals pre-fed the instrumental outcome ($r_{(22)} = 0.30$, $p=0.15$) and if anything there was a non-significant trends towards greater variation in reward rates during the final session in animals that subsequently displayed insensitivity to outcome devaluation. Thus the current data provide little or no evidence to support the assertion that reward rates and changes in the experienced reward rate are critical determinants of the development S-R habits.

However, there was some evidence that local response rates did vary as a function of the amount of training animals had received, with lower variation in the High training group relative to the Low and Medium groups (mean variation in response rate Low group = 4.5; Medium group = 5.0; High group = 4.1), although statistically there was no evidence to support this observation ($F_{(2,42)} = 1.95$, $p=0.31$). The variation (indexed by the standard deviation of the local response rate in the final training session) in response rates

on the final day of acquisition and extinction performance for the animals pre-fed the instrumental outcome are plotted on Figure 2.3.3. The suggestion from this figure is that animals that experienced the greatest variation in response rates across the final session of training showed a more marked sensitivity to the change in reward value after pre-feeding with the instrumental outcome. Conversely, animals that had experienced more stable response rates in the training session preceding the extinction test appeared to be less sensitive to devaluation by specific satiety.

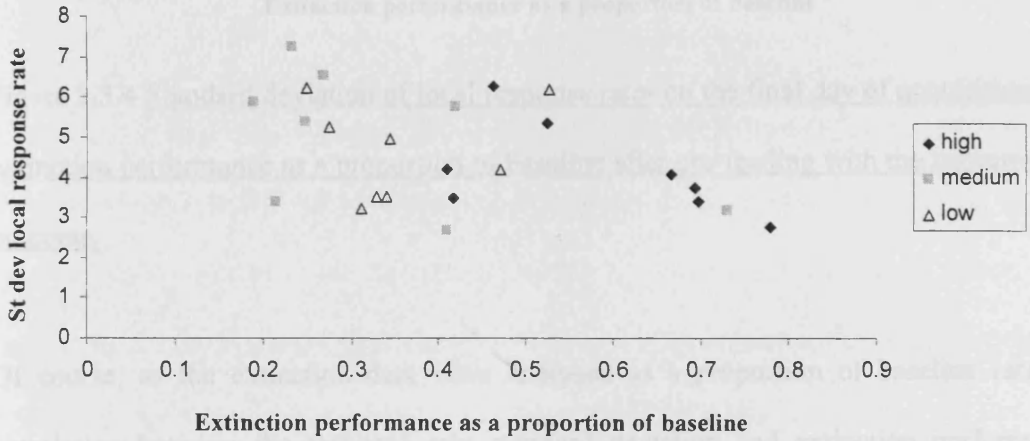

Figure 2.3.3 Standard deviation of local response rates on the final day of acquisition and extinction performance as a proportion of baseline after pre-feeding with the instrumental outcome.

Inspection of Figure 2.3.4 confirms that there was a negative relationship between variation in response rates during the final session of instrumental training and sensitivity to goal-value in the subsequent extinction test ($r_{(22)} = -0.45$, $p<0.05$).

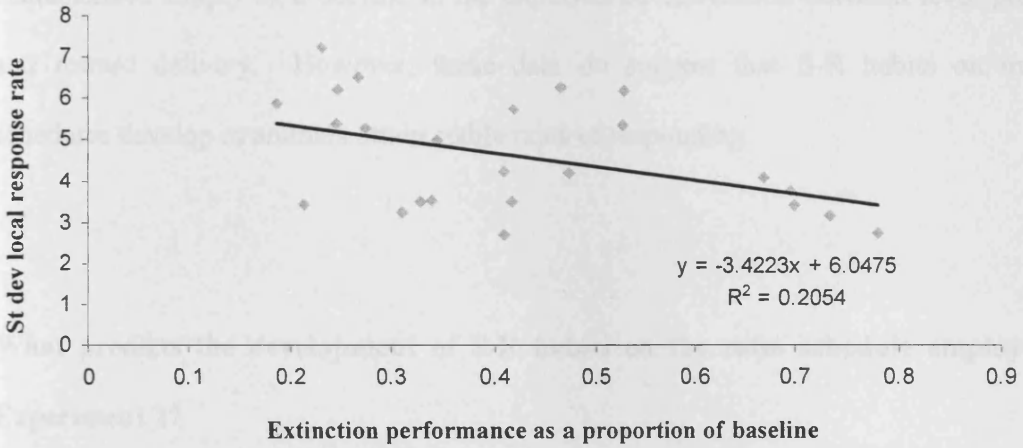

Figure 2.3.4 Standard deviation of local response rates on the final day of acquisition and extinction performance as a proportion of baseline after pre-feeding with the instrumental outcome.

Of course, as the extinction data were analysed as a proportion of baseline rates, a correlation between the response rate standard deviation and extinction performance would to some extent be expected. In order to account for this possibility, a partial correlation between these factors controlling for baseline responding was calculated. This revealed a statistically significant negative correlation between these factors ($r_{(21)}$ = -0.44, p<0.05) confirming that the degree of variation in response rates during the final session of instrumental acquisition predicted the sensitivity to outcome devaluation observed at test. Significantly, variation in response rates in the final session bore little or no relationship to variation in reward rates ($r_{(22)}$ = -0.12, p=0.59). Thus animals that produced more stable response rates did not necessarily experience less variation in

reward rates. As such, the finding in the current experiment that reduced variability in response rates was predictive of sensitivity to outcome devaluation can not be characterised simply as a decline in the experienced correlation between lever pressing and reward delivery. However, these data do suggest that S-R habits on interval schedules develop as animals attain stable rates of responding.

What predicts the development of S-R habits on the ratio schedule employed in Experiment 2?

To date, little empirical or theoretical work has addressed the issue of ratio schedules and the development of S-R habits. Previous research has shown that training on interval and ratio schedules can promote differential sensitivity to changes in reward value; after equivalent amounts of training (2 sessions earning 30 reinforcers per session), a response trained on a ratio schedule, but not on an interval schedule, remained goal-sensitive (Dickinson et al., 1983). It has been suggested that this effect arises because ratio schedules always maintain correlated variation in response and reward rates (Dickinson, 1985).

As discussed previously, ratio schedules establish a markedly different behaviour-reward function compared to interval schedules. Figure 2.4.1 displays the response and reward rates for the final day of instrumental training on the RR8 schedule employed in Experiment 2. The linear relationship between lever press rates and reward rates on ratio

schedules is evident from this figure. As lever press rates rose, there was a concomitant increase in the reward rates ($r_{(46)} = 0.91, p<0.05$).

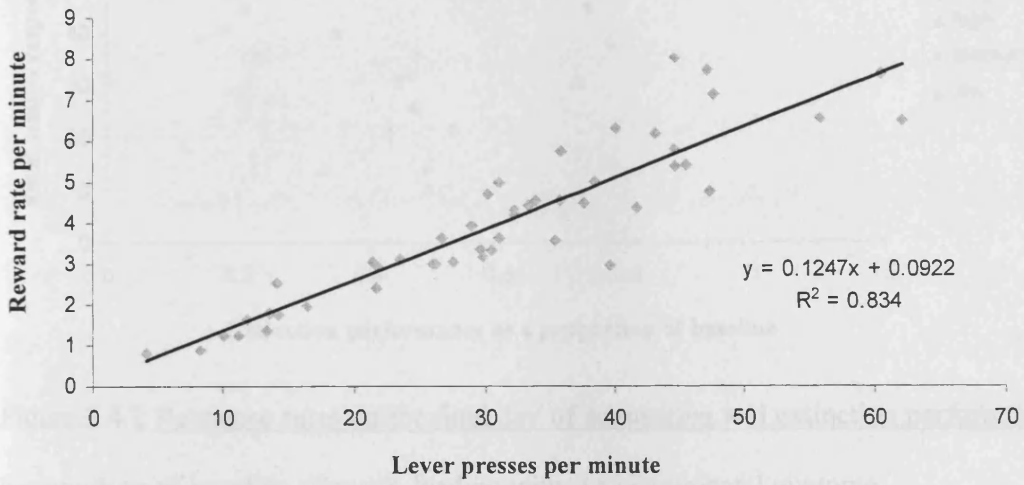

Figure 2.4.1 Reward-response rate function on the final day of RR training in Experiment 2.

As there were significant differences in baseline rates (see Figure 2.2.1 above), it follows that animals' performance as measured on Figure 2.4.1 varied according to the amount of training received. As sensitivity to outcome devaluation also varied as a function of the amount of training received, baseline rates could, therefore, be expected to predict performance in the extinction test following pre-feeding with the instrumental outcome. However as is clear from Figure 2.4.2, there was no relationship between global response rates during the final session of instrumental training and subsequent sensitivity to outcome devaluation. This was confirmed statistically ($r_{(22)} = -0.15$, p=0.47). Similarly, global reward rates did not correlate with extinction performance ($r_{(22)} = -0.09$, p=0.67).

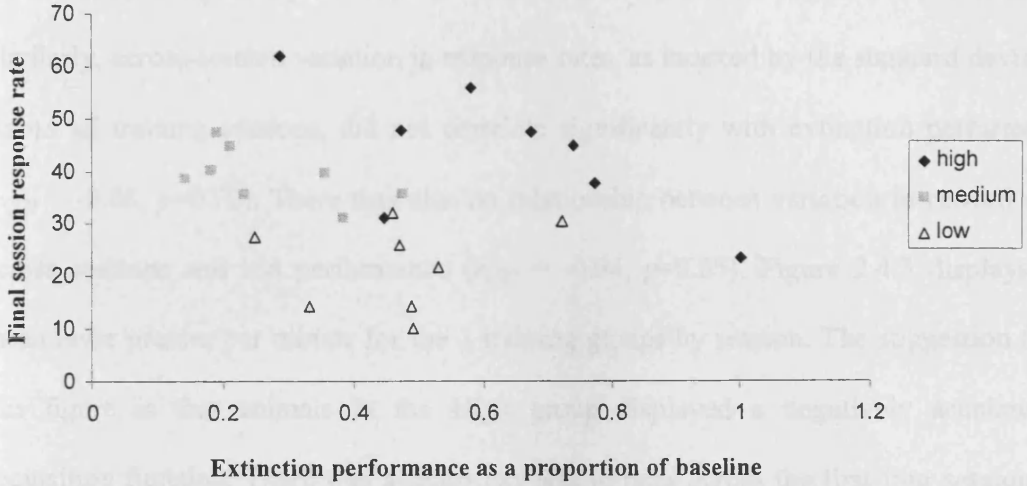

Figure 2.4.2 Response rates on the final day of acquisition and extinction performance as a proportion of baseline after pre-feeding with the instrumental outcome.

Analysis based on global rates may mask variation in local rates. As an index of variation in local rates, the standard deviations for the local response and reward rates (1 minute bins) across the final session were calculated. However, there was no evidence to suggest that variation in response rates during the final session predicted subsequent extinction performance (mean variation in response rates (±SD) Low group = 7.1 (±2.5); Medium group = 9.8 (±4.0); High group = 14.9 (±6.7)). ($r_{(22)}$ = -0.05, p=0.82). Furthermore, variation in reward rates during the final session of instrumental acquisition bore no significant relationship to sensitivity to outcome devaluation (mean variation in reward rates (±SD) Low group = 1.3 (±0.43); Medium group = 1.4 (±0.28); High group = 2.2 (±0.6)). ($r_{(22)}$ = 0.28, p=0.18).

Thus in contrast to the effect seen after training on an interval schedule in Experiment 1, within session variability in response rates did not relate to extinction performance. Similarly, across-session variation in response rates, as indexed by the standard deviation across all training sessions, did not correlate significantly with extinction performance ($r_{(22)}$ = -0.08, p=0.70). There was also no relationship between variation in reward rates across sessions and test performance ($r_{(22)}$ = -0.04, p=0.85). Figure 2.4.3 displays the mean lever presses per minute for the 3 training groups by session. The suggestion from this figure is that animals in the High group displayed a negatively accelerating acquisition function. There was a sharp increase in rates across the first four sessions of instrumental training but thereafter rates began to stabilise and eventually reached a plateau as animals attained some asymptote level of performance. On the other hand, animals in the Low and Medium group still experienced marked session by session increments in response rates.

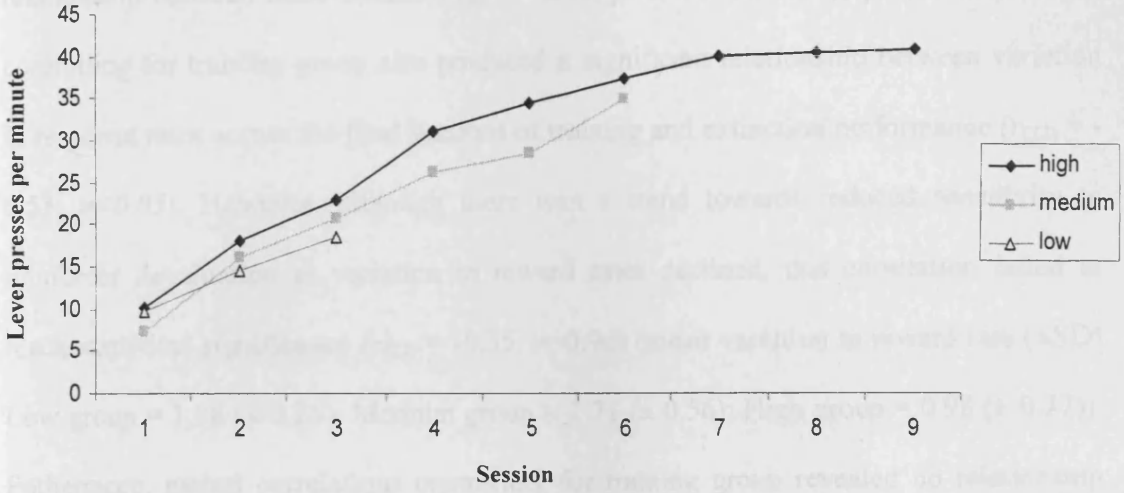

Figure 2.4.3 Mean response rates by session for the Low, Medium and High training groups.

As on average the animals in the High group showed no sensitivity to outcome value at test, it would appear that S-R habits developed as a function of the decline in session by session changes in response rates. This was confirmed statistically. Variation in response rates across the final sessions of instrumental training were explored. Significant effect s were only found when sessions 4-9 High group, 1-6 Medium group and 1-3 Low group were analysed. The variation in response rates across these sessions, as indexed by the standard deviation, is plotted in Figure 2.4.4 against extinction performance after pre-feeding with the instrumental outcome. It is clear from this figure that there was a negative relationship between variation in response rates across the final sessions of training and subsequent sensitivity to outcome devaluation. Statistical analysis revealed a significant negative correlation between these factors ($r_{(22)}$ = -0.53, $p<0.05$). Partial correlations controlling for baseline performance revealed an equally robust negative

relationship between these factors ($r_{(22)}$ = -0.51, p<0.05). Similarly, partial correlations controlling for training group also produced a significant relationship between variation in response rates across the final sessions of training and extinction performance ($r_{(22)}$ = -0.53, $p<0.05$). However, although there was a trend towards reduced sensitivity to reinforcer devaluation as variation in reward rates declined, this correlation failed to reach statistical significance ($r_{(22)}$ = -0.35, p=0.96) (mean variation in reward rate (±SD) Low group = 1.38 (± 0.25); Medium group = 1.71 (± 0.56); High group = 0.98 (± 0.37)). Futhermore, parital correlations controlling for training group revealed no relationship between variation in reward rates and sensitivity to outcome devaluation at test ($r_{(22)}$ = -0.28, p=0.2).

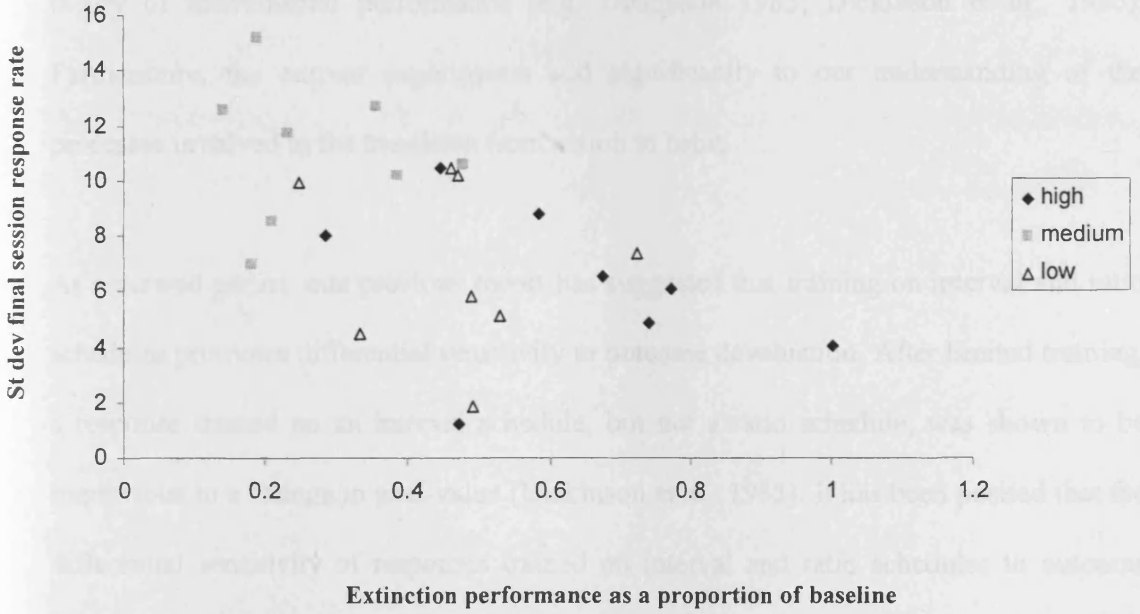

Figure 2.4.4 Standard deviation of response rates across the final sessions of instrumental acquisition and extinction performance as a proportion of baseline after pre-feeding with the instrumental outcome.

Discussion

The simple experiments reported in this chapter examined the sensitivity of an instrumental response, trained either on an interval or ratio schedule, to outcome devaluation after varying amounts of training. The results demonstrated that overtraining of an instrumental response engenders behaviour that is resistant to reinforcer devaluation. There is now a voluminous literature indicting that over the course of instrumental training, response control is ceded from its consequences to its antecedents (e.g. Adams, 1982; Dickinson et al., 1983; Dickinson et al., 1995; Dickinson et al., 1998; Killcross and Coutureau, 2003; Coutureau and Killcross, 2003; Yin et al., 2004; Yin et al., 2006). As such, the current findings provide yet further evidence for a dual-process theory of instrumental performance (e.g. Dickinson 1985; Dickinson et al., 1995). Furthermore, the current experiments add significantly to our understanding of the processes involved in the transition from action to habit.

As reviewed earlier, one previous report has suggested that training on interval and ratio schedules promotes differential sensitivity to outcome devaluation. After limited training, a response trained on an interval schedule, but not a ratio schedule, was shown to be impervious to a change in goal-value (Dickinson et al., 1983). It has been posited that the differential sensitivity of responses trained on interval and ratio schedules to outcome devaluation is the product of the different behaviour-reward functions established by these schedules. In the current experiments, the different behaviour-reward function of these schedules was evidenced by differences in baseline response rates. Animals trained

on ratio schedules pressed the lever at markedly higher rates than the animals trained on interval schedules after equivalent amounts of training. This is a ubiquitous effect that has been shown to persist even if the schedules are matched in terms of reward rates and the temporal patterning of reinforcement (e.g. Catania, Matthews, Silverman and Yohalem, 1977; Dawson and Dickinson, 1990; Baum, 1993; Cole, 1994). This has led to suggestions that the difference in response rates arises from the different feedback functions of these schedules (e.g. Baum, 1981; Baum, 1993). Yet despite the different behaviour-reward functions established by these schedules, training on a ratio schedule did not appear to attenuate the onset of S-R habits. Similarly, there was no evidence that the interval schedule used in Experiment 1 promoted the more rapid onset of S-R habits compared to ratio schedules. Sensitivity to outcome devaluation observed after 120, 240 and 360 response-reward pairings was comparable across the two experiments. Thus it would appear unlikely that habits develop simply as a function of the relationship between behaviour and reward. Indeed, there was no evidence in the current experiments to suggest that an animal's performance as measured on the behaviour-reward curve predicted sensitivity to outcome devaluation.

However, the experiments reported here did reveal variation in response rates as a critical determinant of sensitivity to outcome devaluation. In Experiment 1, a negative correlation between local variation in response rates across the final session and test performance was found. On the ratio schedule employed in Experiment 2, responding became independent of the current value of the reinforcer as between-session variation in response rates declined. The finding that sensitivity to outcome devaluation was predicted

by within-session variation on interval schedules but by between-session variation on ratio schedules is perhaps not surprising in view of the finding that ratio schedules tend to produce overall higher rates of responding. As demonstrated in Experiment 1, training on interval schedules tends to produce a negative accelerating acquisition function as animals very rapidly reach near-asymptote levels of performance. Consequently, rates do not vary significantly across sessions. Even though overall rates were equivalent across the three training groups, within-session variability in response rates did decline over the course of training. The reverse holds for training on ratio schedules: animals tend to experience relatively large session by session changes in response rates early in training but eventually the acquisition curve plateaus as performance reaches an as asymptote. Furthermore, as the rate at which response rates stabilize will vary from animal to animal (and hence experiment to experiment), the amount of training required for habitual responding to emerge will also vary. This may explain the discrepancy between the findings reported here and previous work (Dickinson et al., 1983).

Although the exact measure that predicted test performance differed across the two experiments (within-session variability on interval schedules and between-session variability on ratio schedules), these results broadly suggest that instrumental responding becomes independent of reinforcer value as animals achieve a stable rate of responding. This finding is generally consistent with previous suggestions that habits arise due to a decline in the variation in behaviour (Dickinson, 1985). There was, however, no evidence that variation in reward rates correlated with test performance. This suggests that once a stable level of performance has been established, animals are no longer sensitive to

variation in local reward rates or the molecular relationship between behaviour and reward. Rather, a certain level of performance becomes associated with a reliable and well-predicted reward rate and behaviour is relatively unaffected by minor fluctuations in reward rates. Recent neurophysiological evidence suggests a neural correlate for the observation that as response rates ceased to vary, there was a marked decline in sensitivity to outcome devaluation. It has been shown that over the course of habit learning neural activity within the sensorimotor striatum undergoes dynamic reorganisation but critically, once behavioural asymptote is reached, neuronal activity remains stable or declines during subsequent task. Neuronal activity no longer marks individual task elements but rather the beginning and end of the procedure is emphasised (Carelli, Wolske and West, 1997; Jog et al., 1999; Barnes et al., 2005). Arguably, the more fundamental question is perhaps not why habits develop as behaviour becomes less variable but rather, how this process occurs. This issue will be dealt with in Chapter 6.

Most previous investigations into the development of S-R habits have only compared the performance of under- and over-trained groups. The inclusion in the current experiments of a Medium group allowed greater analysis of the processes involved in the development of S-R habits. One of the striking aspects of the data from the Medium groups was that, even after 6 sessions of training on an interval schedule, animals in Experiment 1 were still sensitive to changes in reward value. This provides yet further evidence that training on interval schedules does not necessarily promote a more rapid transition from action to habit than ratio schedules. Furthermore, the finding in both experiments of overall lower responding in the Medium groups in the extinction tests relative to the other training

groups has implications for understanding how R-O and S-R mechanisms interact at a psychological level. It is known that, while reinforcer devaluation leads to a selective reduction in responding controlled by goal-directed actions, devaluation manipulations, such as pre-feeding to satiety, produce a *general* suppression in responding due to the impact of reduced motivation on S-R associations (Dickinson and Balleine, 1994; Dickinson et al., 1995). Thus, the relatively low rates of responding seen in the Medium groups at test might be taken as evidence of both the selective impact of the devaluation procedure on R-O associations, as well as a general suppression in performance as a result of the non-selective effects of the motivational shift on S-R associations. Such findings render interpretation of the transition from action to habit as reflecting changes in an animal's knowledge about the A-O contingency problematic. Rather, they suggest that the transition from action to habit involves active competition between the goal-directed and habit systems: as instrumental training proceeds, A-O associations persist but are gradually overridden by S-R mechanisms. Indeed, the finding here that A-O and S-R processes can be engaged concurrently is consistent with the results from recent investigations into the neural substrates of instrumental conditioning. These studies have revealed that disruption to the habit system abolishes response control by S-R habits and renders instrumental performance goal-directed (e.g. Killcross and Coutureau, 2003; Coutureau and Killcross, 2003; Yin et al., 2004; Yin et al., 2006).

A further notable feature of the current experiments was that magazine approach behaviour remained sensitive to outcome devaluation even after extensive training. This dissociation may reflect the greater control of magazine approach by Pavlovian rather

than instrumental contingencies. Indeed, there is very good evidence that Pavlovian conditioned responses, such as magazine entry behaviour, remain sensitive to outcome devaluation even after extensive training (Holland, 1998; Holland, 2005). This difference may be related to the finding that the impact of motivational manipulations on instrumental performance, but not Pavlovian responses, is mediated by incentive learning (Balleine, 1992; Dickinson and Balleine, 1994). For example, Balleine (1992) demonstrated that magazine approach behaviour is immediately sensitive to changes in an animal's motivational state. Alternatively, as the animals in the current experiments were required to push back a panel to access rewards in the magazine recess, it could be claimed that animals had to perform a heterogeneous chain of responses to obtain rewards. In this respect, it is significant that research has shown that actions proximal, but not distal, to rewards are sensitive to motivational manipulations (Balleine, Garner, Gonzalez and Dickinson, 1995; Killcross and Blundell, 2001; Daw et al., 2005). The current findings are therefore consistent with previous work suggesting that Pavlovian and instrumental processes can be mediated by different psychological and neural mechanisms (e.g. Dickinson, Smith and Mirenowicz, 2000; Corbit et al., 2001; Dayan and Balleine, 2002; Corbit and Balleine, 2003). Significantly for the current discussion, the dissociation between lever press performance and magazine approach behaviour provides unequivocal evidence that the pre-feeding procedure successfully devalued the outcome in all groups.

In summary, the experiments reported here add significantly to our understanding of the psychological processes involved in the transition from action to habit. It has been

demonstrated that the reinforcement schedule employed during training does not necessarily influence the rate at which habits develop. Significantly, the degree of variation in response rates has been identified as a critical determinant of sensitivity to outcome devaluation. Furthermore, the current findings provide preliminary evidence that R-O and S-R processes can operate concurrently.

CHAPTER 3

Experiments 3-7: The effects of changes in the schedule of reinforcement and reward magnitude on the sensitivity of an overtrained instrumental response to outcome devaluation.

The previous chapter identified variability in behaviour as a fundamental determinant of the transition from action to habit. These results suggested that habits developed as animals established a stable response rate. The experiments also provided preliminary behavioural evidence to suggest that A-O associations are not lost over the course of training but are gradually dominated by S-R associations. The experiments reported here sought to explore these effects further by analysing the impact of behavioural manipulations on the performance of instrumental responses under the control of S-R habits. To date, little work has addressed the issue of what factors affect the performance of habitual S-R responses. For example, there is some evidence that contextual manipulations can disrupt S-R habits (McDonald, King and Hong,, 2001). However, there have been no empirical investigations of factors that influence the sensitivity of overtrained instrumental responses to outcome devaluation. Similarly, little or no behavioural data have been reported to support evidence from lesion studies that even after extensive training, response control can switch from habitual to goal-directed control (e.g. Coutureau and Killcross, 2003; Yin et al., 2006).

Experiments 3-5 explored the effect of changes in the schedule of reinforcement and reward magnitude on an overtrained instrumental response. Initially, animals received sufficient training (360 response-reward pairings, over 9 sessions) before experiencing either an upward or downward shift in these factors. In Experiment 6, animals were trained on a reinforcement schedule that changed every two minutes. Animals were then given a test of sensitivity to outcome devaluation by specific satiety. The effect of these manipulations was to disrupt both response and reward rates. If responding at test is guided by outcome expectancy, it would suggest not only that perturbations to these factors can restore goal-directed responding but also that stable response and reward rates are critical to both the development and expression of S-R habits. Finally, Experiment 7 explored the suggestion that extended pre-training exposure to the reinforcer facilitates the development of S-R habits (Adams, 1982).

Experiments 3 and 4. The effect of an up- or downward shift in the schedule of reinforcement on the sensitivity of an overtrained instrumental response to outcome devaluation after training on interval (Experiments 3a and 3b) and ratio (Experiment 4) schedules.

Method and Materials

Subjects

Male Lister hooded rats were used in these experiments (Experiment 3a n = 48; Experiment 3b n = 32; Experiment 4 n = 48; Harlan UK Ltd., Bicester, Oxon, UK). At the start of behavioural test animals weighted between 243 g and 339 g. Animal care and husbandry were as described elsewhere.

Behavioural apparatus

Eight operant chambers (30 cm wide by 24 cm deep by 21 cm high; supplied by Med Associates Inc., St Albans, VT), housed in sound attenuating chambers and arranged in a 2 × 4 array, were used. Each chamber consisted of three aluminium walls and ceiling, with a Perspex door serving as the fourth wall. Each chamber had a floor constructed of 19 stainless steel rods (4.8 mm in diameter, spaced 1.6 cm apart). The chambers were illuminated by a 3-W houselight located at the top centre of the left wall. Food pellets (45 mg; Formula A/I, P.J. Noyes, Lancaster, NH) were delivered into a recessed magazine located in the right wall of each chamber. 15% (w/v) sucrose solution was delivered via a dipper (vol = 0.1 ml) into the same magazine. Access to the magazine could be determined by means of infrared detectors mounted across the mouth of the recess. A flat-panel retractable lever could be inserted to the left and right of the magazine. A light located in the roof served as the houselight. A computer equipped with MED-PC® software (Med Associates Inc.) controlled the operant chambers and recorded the data.

Behavioural Training

Magazine training. Magazine training proceeded as described previously.

Lever press training. In order to produce behaviour that would normally be insensitive to outcome devaluation, all animals in Experiments 3a and 3b received identical training to the High group in Experiment 1. The rewards were either food pellets (Noyes) or sucrose solution (15% w.v.). Thus after initial acquisition on a CRF schedule, animals earned 40 rewards a session for 9 sessions on a RI30 schedule of reinforcement. Thereafter all the animals received a further single session again earning 40 reinforcers albeit on different reinforcement schedules. In this session in Experiment 3a one third of the animals earned the reinforcers on the same RI30 reinforcement schedule (group RI30), one third was switched to a richer reinforcement schedule (group RI9) and lever pressing for the final 16 animals was now reinforced on a leaner RI100 schedule (group RI100). Experiment 3b proceeded in exactly the same manner except that in the 10^{th} session half the animals continued to earn rewards on the RI30 schedule (group RI30) and the other half was switched to the leaner RI100 schedule (group RI100). In Experiment 4, after CRF training animals were trained for 9 sessions on a RR9 schedule earning 40 reinforcers per session and 360 in total (sufficient to produce habitual responding as demonstrated in Experiment 2). In a subsequent session, 16 animals again earned 40 reinforcers on a RR9 schedule (group RR9) 16 now earned 40 rewards on a richer reinforcement schedule (group RR3) and 16 on a leaner schedule (group RR30).

Devaluation by specific satiety (Experiments 3a and 4)

Following lever press training, all animals received 1 session of specific-satiety devaluation extinction testing. As there is some evidence that the reduced sensitivity of

overtrained instrumental responses to outcome devaluation can be masked by the excitatory potential of contextual cues, the devaluation procedures were conducted in the conditioning chambers (Coutureau and Killcross, 2003). On the day after final day of instrumental training, each rat was placed in its experimental chamber and 40 rewards were delivered on a RT 60-s schedule. Half the animals received presentations of the sucrose solution and the other the Noyes pellets. Thus half the animals were prefed the outcome earned during instrumental training (devalued group) and the other half an alternative reinforcer (non-devalued group). The house light was illuminated throughout this session but the lever was not available. Immediately after the cessation of this outcome devaluation session, the animals were removed from the chambers and any uneaten rewards were removed from the magazine tray. The animals were then transferred back to the conditioning chambers and received a 10 minute extinction test during which lever press responding was measured in the absence of reward delivery.

Devaluation by Lithium chloride (Experiment 3b)

In Experiment 3b, the reinforcer was devalued using conditioned taste aversion. After the final session of instrumental training, animals received three days of devaluation with lithium chloride (LiCl). On each day the rats were placed in the operant chambers and were given 40 free presentations of the instrumental outcome on an RT 60-sec schedule. Immediately after the cessation of each session, the devalued group received a 0.15M, 10 ml/kg (*i.p.*) injection of LiCl solution (Sigma Chemical Co., Poole, UK) and the non-devalued group an injection of the equivalent volume of saline. 24 hours after the final

session of taste aversion training, animals' sensitivity to outcome devaluation was assessed in an extinction test conducted as described above.

Data analysis

Statistical analysis was performed using analysis of variance (ANOVA) with between subject factors of the reinforcement schedule used on the final day of instrumental training (same, change up or change down) and devaluation by either specific satiety or LiCl (devalued versus non-devalued).

Results

Experiment 3a: The effect of an up- or downward shift in the schedule of reinforcement on sensitivity of an overtrained instrumental response to devaluation by specific satiety after training on interval schedules.

Instrumental training – RI30 training. All animals acquired the instrumental response. By the 9th day of RI30 training, there was an effect of training group ($F_{(2,42)}$ = 3.71, $p<0.05$) reflecting overall higher rates of responding in the RI9 group (Mean lever presses per minute (± SD) RI30 group = 14.6 (±2.9); RI9 group = 18.2 (± 3.7); RI100 group = 16.327 (± 4.216)). Post-hoc Tukey tests revealed that that the RI9 group responded at higher rates than the RI30 ($p<0.05$). As group allocation was made prior to training and until this point all animals had been trained on the same RI30 reinforcement

schedule, this difference simply reflects baseline differences in response rates rather than any experimental manipulation. Significantly, there was no effect of intended devaluation or an interaction between this factor and the reinforcement schedule to be used in 10^{th} session (both Fs<1). Moreover, there were no differences in magazine entry behaviour between any of the groups (all Fs<1).

Instrumental training – change in reinforcement schedule. On the 10^{th} day of instrumental training the two experimental groups were exposed to a different reinforcement schedule (either RI9 or RI100) and the Same group continued to earn rewards on the RI30 schedule. In terms of lever presses per minute (see Figure 3.1.1), this manipulation was without effect as there was no effect of reinforcement schedule, intended devaluation or any interaction (all Fs <1). Moreover, total lever presses differed considerably between the three groups ($F_{(2,42)} = 65.44$, $p<0.05$) (Mean total lever presses (± SD): RI30 group = 305.5 (± 87.9); RI9 group = 110.9 (± 21.9); RI100 group = 869.6 (± 322.1)). Post-hoc Tukey tests confirmed that the total number of lever presses in RI9 group was significantly lower than both the RI30 group ($p<0.05$) and the RI100 group ($p<0.05$) and that RI30 and RI100 also differed ($p<0.05$). Again, there were no effects of devaluation or any interaction (both Fs<1). Critically, as is clear from Figure 3.1.1, response rates were disrupted relative to baseline levels in the two groups that experienced a change in reinforcement schedule. ANOVA confirmed that rates did differ across the 2 sessions ($F_{(1,42)} = 105.97$, $p<0.05$). The decline in rates was most pronounced in the two experimental groups as there was a session x schedule interaction ($F_{(2,42)} = 7.55$, $p<0.05$). Simple effects confirmed that there was a significant fall in rates relative to

baseline in the animals exposed to the RI9 schedule ($F_{(1,14)}$ = 63.83, p<0.05) and the RI100 schedule ($F_{(1,14)}$ = 49.29, p<0.05).

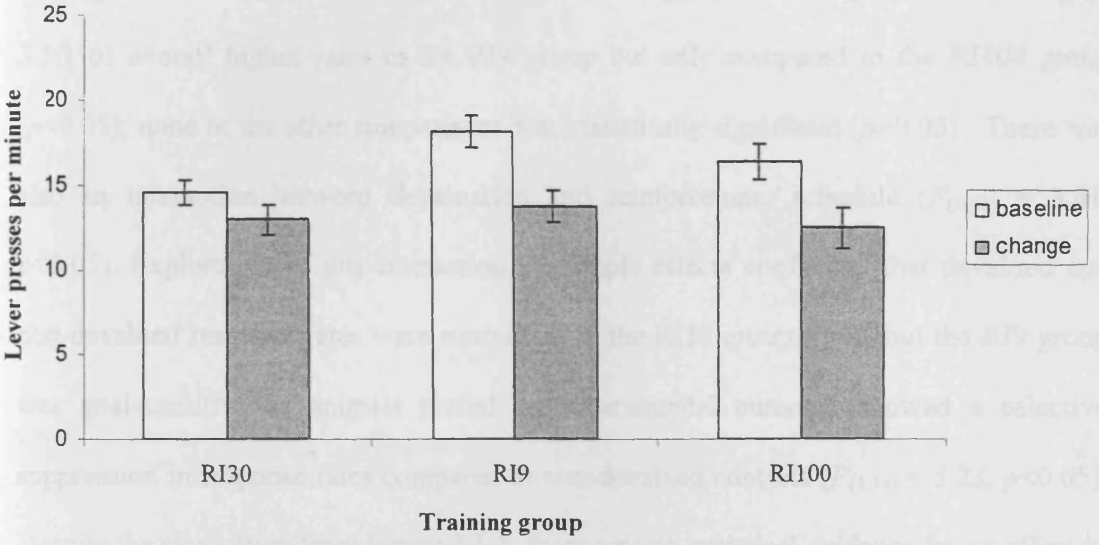

Figure 3.1.1. Mean lever presses per minute (±SEM) before (white bars) and after(grey bars) the change in reinforcement schedule.

Extinction test – lever press performance. The mean lever presses per minute as a proportion of response rates during the session 10 (i.e. when animals experienced a change in reinforcement schedule) are presented in Figure 3.1.2. As expected, inspection of this figure reveals that instrumental performance in the animals consistently trained on a RI30 schedule was habitual as there was no selective effect of the pre-feeding procedure. However, a switch to a richer reinforcement schedule appeared to restore goal-sensitivity as the devalued RI9 group showed a clear suppression in lever press rates compared to the non-devalued RI9 animals. Conversely, the direction of the devaluation effect was reversed in the RI100 animals so that rates were higher in the devalued group

relative to the non-devalued group. This is description of the data was partially supported by ANOVA which yielded no effect of devaluation ($F<1$) but an effect of schedule ($F_{(2,42)} = 3.8$, $p<0.05$). Post-hoc Tukey tests of this effect supported the impression from Figure 3.1.1 of overall higher rates in the RI9 group but only compared to the RI100 group ($p<0.05$); none of the other comparisons was statistically significant ($p>0.05$). There was also an interaction between devaluation and reinforcement schedule ($F_{(2,42)} = 3.81$, $p<0.05$). Exploration of this interaction by simple effects confirmed that devalued and non-devalued response rates were equivalent in the RI30 group ($F<1$) but the RI9 group was goal-sensitive as animals prefed the instrumental outcome showed a selective suppression in response rates compared to non-devalued controls ($F_{(1,14)} = 5.23$, $p<0.05$). Despite the suggestion from Figure 3.1.1, there was no statistical evidence for an effect of the devaluation procedure in the RI100 ($F_{(1,14)} = 2.56$, p=0.13).

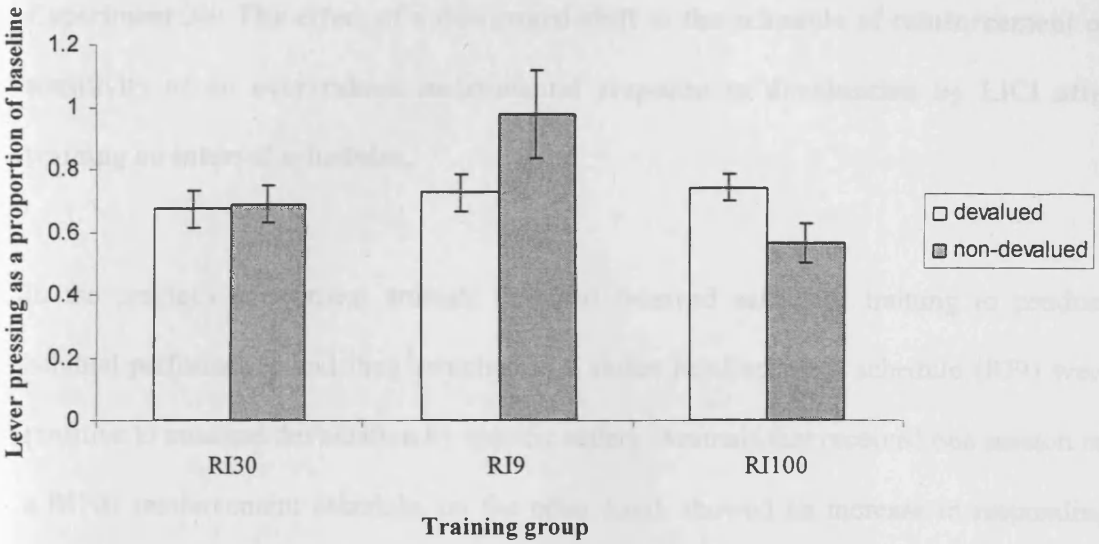

Figure 3.1.2 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after pre-feeding with the instrumental outcome (devalued - white bars) or an alternative reinforcer (non-devalued - grey bars).

Extinction test – magazine entry performance. Despite the differential sensitivity to outcome devaluation of lever pressing, magazine approach behaviour remained sensitive to devaluation by specific satiety in all groups. ANOVA confirmed an effect of devaluation ($F_{(1,42)}$ = 7.17, p<0.05) but no effect of reinforcement schedule and no interaction between these factors (both Fs<1) (Mean magazine entries per minute as a proportion of baseline (± SD): devalued RI30 group = 0.34 (± 0.24); non-devalued RI30 group = 0.65 (± 0.31); devalued RI9 group = 0.32 (± 0.2); non-devalued RI9 group = 0.49 (± 0.34); devalued RI100 group = 0.42 (± 0.21); non-devalued RI100 group = 0.61 (± 0.37).

Experiment 3b: The effect of a downward shift in the schedule of reinforcement on sensitivity of an overtrained instrumental response to devaluation by LiCl after training on interval schedules.

In the previous experiment animals that had received sufficient training to produce habitual performance and then switched to a richer reinforcement schedule (RI9) were sensitive to outcome devaluation by specific satiety. Animals that received one session on a RI100 reinforcement schedule, on the other hand, showed an increase in responding after pre-feeding with the instrumental outcome. The switch to such a lean schedule of reinforcement may have produced conditions akin to extinction and hence the increase in responding relative to non-devalued performance after pre-feeding with the instrumental outcome could be interpreted as a reinstatement effect, and as such may be taken as evidence of goal-sensitivity. Statistically, however, the increase in devalued performance relative to non-devalued performance was not significant. Experiment 3b aimed to provide further evidence that downward shifts in the schedule of reinforcement reinstate goal-directed responding to over-trained instrumental responses by again examining the effects of reinforcer devaluation on instrumental performance after 9 sessions of RI30 training followed by a single session of RI100 training. However, by employing a more potent method of reinforcer devaluation (LiCl-induced gastric-malaise), unequivocal evidence of outcome expectancy could be sought.

Instrumental training – RI30 training. All animals acquired the instrumental response and by the 9th day of RI30 training there was no effect of intended devaluation,

reinforcement schedule nor an interaction between these factors (all Fs<1) (Mean lever presses per minute (± SD): RI30 group = 18.9 (±4.3); RI100 group = 19.9 (±3.8). Similarly there was no effect of either factor on magazine entry behaviour and no interaction (highest $F_{(1,28)} = 1.84$, p=0.19).

Instrumental training – change in reinforcement schedule. Analysis of lever press rates during the session in which the experimental group was reinforced on a RI100 schedule, revealed a non-significant trend towards lower response rates in the RI100 group ($F_{(1,28)} =$ 3.21, p=0.08) but no effect of intended devaluation (F<1) nor any interaction between group and devaluation ($F_{(1,28)} = 1.4$, p=0.25). Subsequent analysis in terms of the total responses required to earn 40 reinforcers confirmed a highly significant effect of the change in schedule ($F_{(1,28)} = 139.73$, p<0.05) but no effect of devaluation (F<1) nor an interaction ($F_{(1,28)} = 1.49$, p=0.23) (Mean total lever presses (±SD): RI30 group = 400.8 (± 84.1); RI100 group = 1100.5 (±391.5)). Figure 3.1.3 displays the mean lever presses per minute on the final day of RI30 training and on the 10th day of training when the RI100 group experienced the change in reinforcement schedule. It is clear from this figure that this shift in schedule produced a reduction in rates relative to baseline. ANOVA yielded a main effect of session ($F_{(1,28)} = 25.18$, p<0.05) but also an interaction between this factor and reinforcements schedule ($F_{(1,28)} = 7.73$, p<0.05). Simple effects confirmed that while rates across the 2 sessions did not vary in the animals that did not experience a change in reinforcement schedule (F<1), rates were markedly lower in the RI100 group after the change in schedule ($F_{(1,14)} = 30.41$, p<0.05). Furthermore as is clear

from Figure 3.13, response rates in the RI100 group were lower compared to the RI30 group on the 10th day of training ($F_{(1,14)}$ 7.79, p<0.05).

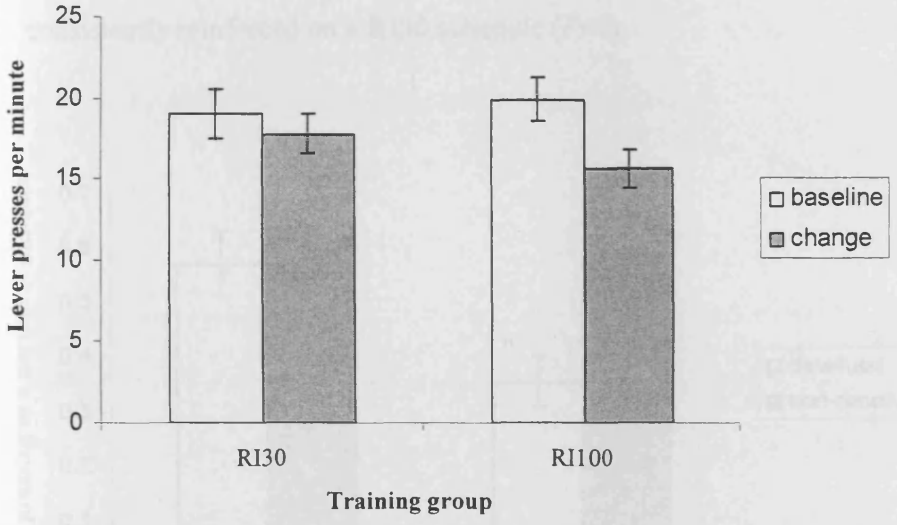

Figure 3.1.3 Mean lever presses per minute (±SEM) before (white bars) and after (grey bars) the change in reinforcement schedule.

Extinction test – lever press performance. The mean lever presses per minute as a proportion of response rates during session 10 (i.e. session change on Figure 3.1.3.) in Figure 3.1.4. It is clear from this figure that lever press performance in the animals that had consistently been trained on a RI30 schedule was controlled by S-R habits as there was no effect of the devaluation treatment on responding. Conversely, the animals that experienced one session on the leaner RI100 schedule appeared to be goal-directed as evidenced by the marked suppression in lever press rates of the devalued group relative to the non-devalued group. This description of the data was supported statistically by an

effect of schedule ($F_{(1,28)}$ = 7.81, p<0.05), devaluation ($F_{(1,28)}$ = 9.72, p<0.05) and critically an interaction between these two factors ($F_{(1,28)}$ = 4.33, p<0.05). Subsequent simple effects analysis of this interaction confirmed response control by outcome expectancy in the RI100 animals ($F_{(1,14)}$ = 13.56, p<0.05) but not in the animals consistently reinforced on a RI30 schedule (F<1).

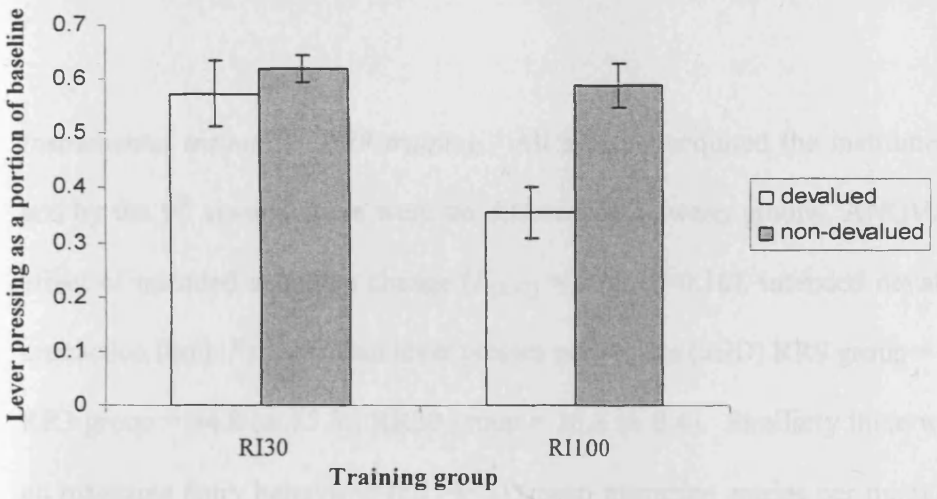

Figure 3.1.4 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after devaluation by LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Extinction test – magazine entry performance. In contrast to lever pressing, magazine approach behaviour in all animals was sensitive to devaluation by LiCl. ANOVA yielded no effect of reinforcement schedule (F<1), a main effect of devaluation ($F_{(1,28)}$ = 9.76, p<0.05) and significantly no interaction between these factors (F<1) (Mean magazine entries per minute as a proportion of baseline (±SD): devalued RI30 group = 0.58

(±0.13); non-devalued RI30 group = 0.86 (±0.29); devalued RI100 group = 0.41 (±0.24) ; non-devalued RI100 group = 0.93 (±0.61)).

Experiment 4: The effect of an up- or downward shift in the schedule of reinforcement on sensitivity of an overtrained instrumental response to devaluation by specific satiety after training on ratio schedules.

Instrumental training – RR9 training. All animals acquired the instrumental response, and by the 9th session there were no differences between groups. ANOVA revealed no effect of intended schedule change ($F_{(2,42)}$ = 1.92, p=0.16), intended devaluation nor an interaction (both Fs<1) (mean lever presses per minute (±SD) RR9 group = 47.5 (± 13.5); RR3 group = 44.8 (± 15.3); RR30 group = 38.8 (± 8.4). Similarly there were no effects on magazine entry behaviour (all Fs<1) (mean magazine entries per minute (±SD): RR9 group = 7.4 (±5.7); RR3 group = 8.0 (±3.0); RR30 group = 5.9 (±6.0)).

Instrumental training – change in reinforcement schedule. On the next day animals experienced no change in reinforcement schedule (group RR9) or were switched to either a richer (RR3) or leaner (RR30) ratio schedule. 7 of the 16 animals failed to maintain responding on the RR30 schedule and consequently the RR30 group was no longer considered for statistical analysis. Analysis of the remaining 2 groups (groups RR9 and RR3) revealed that the switch to a richer reinforcement schedule produced a marked reduction in lever press rates compared to controls that continued to be reinforced on the

RR9 schedule ($F_{(1,28)} = 29.65, p<0.05$) (mean lever presses per minute RR9 group = 45.3; RR3 group = 28.2) but there was no effect of intended devaluation ($F_{(1,28)} = 2.96, p= 0.10$) nor an interaction ($F<1$). Magazine entry behaviour was, however, unaffected by the change in reinforcement schedule as there was no effect of the change ($F_{(1,28)} = 1.67, p= 0.21$), devaluation ($F<1$) nor an interaction ($F_{(1,28)} = 2.69, p= 0.12$). Figure 3.2.1 displays the mean lever presses per minutes for the 9th (baseline) and 10th sessions (change). Comparison of lever press rates across these sessions revealed an effect of session ($F_{(1,28)} = 9.51, p<0.05$). Critically, there was an interaction between this factor and the change in reinforcement schedule ($F_{(1,28)} = 5.93, p<0.05$). As expected, lever press rates in the group that was consistently trained on the RR9 schedule were equivalent across these 2 sessions ($F<1$) but the shift to the RR3 schedule produced a significant decrease in rates ($F_{(1,14)} = 15.23, p<0.05$).

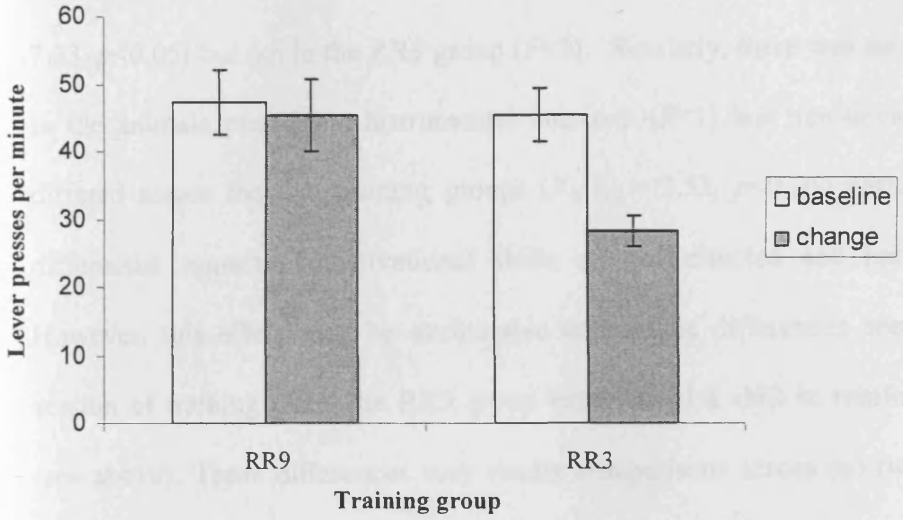

Figure 3.2.1 Mean lever presses per minute (±SEM) before (white bars) and after (grey bars) the change in reinforcement schedule.

Extinction test – lever press performance. Figure 3.2.2 displays the mean lever presses per minute as a proportion of rates during session 10 (i.e. session change) in the extinction test after pre-feeding with either the instrumental outcome (white bars) or an alternative reinforcer (non-devalued - grey bars). The suggestion from this figure is that the change in reinforcement schedule produced behaviour that was sensitive to the current value of the reinforcer as revealed by the selective effect of pre-feeding in these animals. The performance of the animals consistently trained on the RR9 appeared to be independent of this value as there was no effect of the devaluation in the animals prefed the instrumental outcome. This description of the data was supported by statistical analysis. ANOVA yielded no effect of schedule ($F_{(1,28)} = 2.22$, p=0.15) nor an effect of devaluation ($F_{(1,28)} = 2.06$, p=0.16) but an interaction between the two ($F_{(1,28)} = 5.80$,

$p<0.05$). Simple effects confirmed goal-directed response control in the RR3 ($F_{(1,14)} = 7.33$, $p<0.05$) but not in the RR9 group ($F<1$). Similarly, there was no effect of schedule in the animals prefed the instrumental outcome ($F<1$) but non-devalued performance differed across the two training groups ($F_{(1,14)} = 7.53$, $p<0.05$) perhaps reflecting the differential impact of motivational shifts on goal-directed and habitual responding. However, this effect may be attributable to baseline differences seen during the 10th session of training when the RR3 group experienced a shift in reinforcement schedule (see above). These differences may render comparisons across the two training groups problematic. Nevertheless, this can not account for the differential sensitivity of the RR9 and RR3 groups to outcome value seen at test. The critical comparisons are between devalued and non-devalued performance in each of the two training groups. The devaluation effect in the RR3 group arises because of the reduced rates in the RR3 animals prefed the instrumental outcome relative to the non-devalued RR3 group. As both RR3 groups displayed lower response rates during the 10th session of instrumental training compared to both RR9 groups, the selective effect seen in the devalued RR3 group can not be attributed to these baseline differences.

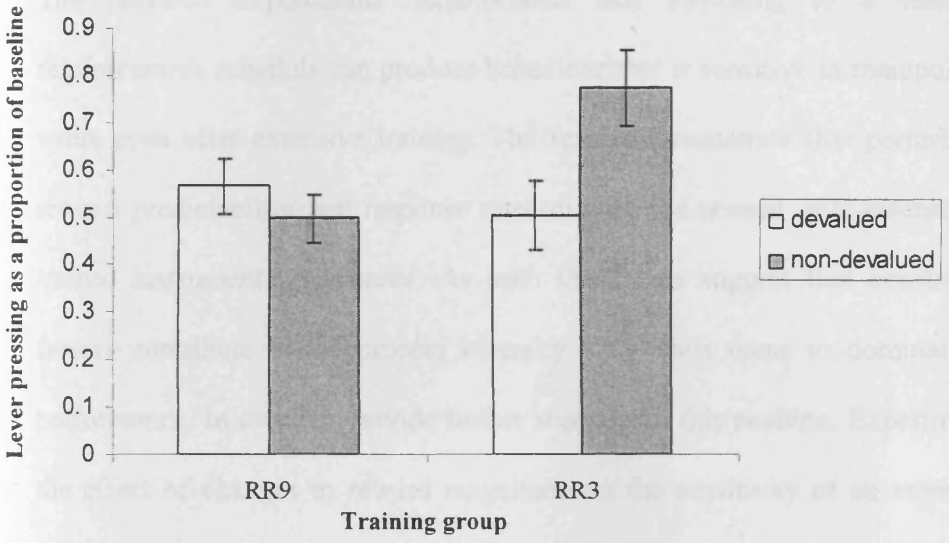

Figure 3.2.2 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after pre-feeding with the instrumental outcome (devalued - white bars) or an alternative reinforcer (non-devalued - grey bars).

Extinction test – magazine entry performance. Magazine approach behaviour was sensitive to goal-value in both training groups as there was an effect of devaluation ($F_{(1,28)}$ = 8.36, p<0.05), no effect of schedule ($F_{(1,28)}$ = 2.89, p=0.1) nor any interaction ($F_{(1,28)}$ = 1.29, p=0.27) (Mean magazine entries per minute as a proportion of baseline (±SD): devalued RR9 group = 0.478 (±0.21); non-devalued RR9 group = 1.0 (±0.64); devalued RR3 group = 0.41 (±0.20); non-devalued RR3 group = 0.63 (±0.49)).

Experiment 5. The effect of changes in reinforcer magnitude on the sensitivity of an overtrained instrumental response to outcome devaluation.

The previous experiments demonstrated that switching to a learner or richer reinforcement schedule can produce behaviour that is sensitive to manipulations in goal-value even after extensive training. The results demonstrate that perturbations to both reward predictability and response rates disrupt the normal goal-insensitivity of over-trained instrumental responses. As such these data suggest that consistency in these factors contribute to the process whereby S-R habits come to dominate instrumental performance. In order to provide further support for this position, Experiment 5 assessed the effect of changes in reward magnitude on the sensitivity of an extensively trained instrumental response to outcome devaluation.

Method and Materials

Subjects

The subjects were 48 experimentally naïve male Lister-hooded rats of mean weight 267g at the beginning of the experiment (range 244 – 302g). Animal care and handling was the same as in previous experiments.

Behavioural apparatus

The experiment was conducted in the same chambers as described for Experiments 3a and 3b.

Behavioural Training

Magazine training. Magazine training proceeded as described previously.

Lever press training. Animals were assigned to one of three groups (Same, Up and Down). During the initial acquisition (2 days of CRF training and 9 days of RI30 training) the Same and Up groups earned 1 pellet and the Down group 3 pellets each time a response was rewarded. Half the animals earned standard Noyes pellets and the other half sucrose pellets (counterbalanced across experimental conditions). As with previous experiments, in each of the 9 RI30 sessions 40 responses were rewarded. In a tenth session, animals in the Same group continued to earn 1 pellet for every rewarded response. The Up group experienced an increase in reward magnitude and now earned 3 pellets for every rewarded response, while the Down group was subject to a reduction in reward magnitude and earned 1 pellet instead of 3, per rewarded response. In this final session the reinforcement schedule (RI30) and the number of response-reward pairings (40) was identical to the previous 9 sessions.

Devaluation by specific satiety

A test of goal-sensitivity by specific satiety was conducted as described for Experiment 3a, but lasted only 8 minutes.

Data analysis

The data were subject to ANOVA with between subject factor of training group (same, up or down) and devaluation (devalued or non-devalued). As the standard deviation was proportional to the mean, logarithmic transformations were performed on the raw scores from the final day of acquisition and the extinction test. Extinction performance is

expressed as a proportion of these transformed data. (Howell, 2002). (Correlation before transformation $r_{(24)} = 0.723$, $p<0.05$; after $r_{(24)} = -0.170$, $p=0.426$).

Results

Instrumental training – RI30 training. All the animals acquired the initial instrumental response but as is clear from Figure 3.3.1 at different rates. The Down group, which earned 3 pellets per rewarded response displayed overall lower rates of responding compared to animals that received only 1 pellet. This is consistent with previous findings that on interval, but not ratio schedules, increases in reward magnitude can attenuate response rates (Reed, 1991). ANOVA supported this observation as there was a main effect of training group ($F_{(2,42)} = 6.61$, $p<0.05$). Post hoc Tukey tests confirmed that responding in the Down group was lower than both the Up ($p<0.05$) and Same groups ($p<0.05$) but these groups did not differ (p=0.992). There was, however, no effect of intended devaluation ($F<1$) nor any interaction ($F<1$). Furthermore, analysis of magazine entry behaviour on 9th day of training revealed no effects (highest $F_{(2,42)} = 1.07$, $p=0.35$) (mean magazine entries per minute (± SD): Same group = 8.8 (±1.7); Up group = 9.9(± 4.1); Down group = 10.4 (±2.6).

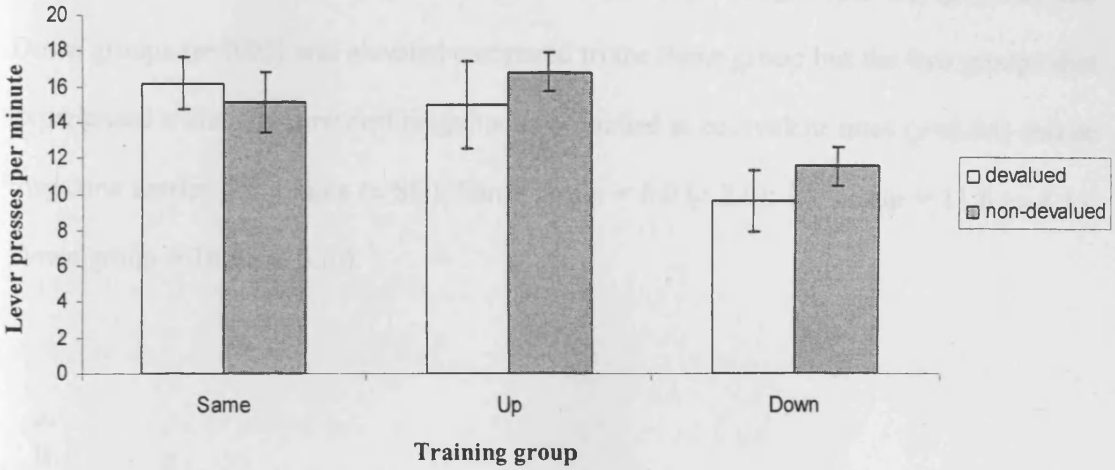

Figure 3.3.1 Mean lever presses per minute (±SEM) on the 9th day of RI30 training for groups earning 1 (Same and Up) or 3 reinforcers (Down).

Instrumental training – change in reinforcer magnitude. In the next session animals experienced a change in reinforcer magnitude so that the Up group now earned 3 pellets and the Down group 1 per rewarded response. Controls experienced no change (Same group). As is clear from *F*igure 3.3.2 the change in reinforcer magnitude led to reduced responding in the Up group and conversely an increase in the Down group. This description of the data was confirmed statistically by ANOVA which yielded an effect of training group ($F_{(2,42)}$ = 10.44, p<0.05), no effect of intended devaluation ($F_{(1,42)}$ = 2.2, p=0.15) and no interaction between these factors (F<1). Tukey tests confirmed that the Up group responded at lower rates compared to both the Same (p<0.05) and the Down group (p<0.05) but rates in these groups did not differ (p=0.16). Similarly, the change in reinforcer magnitude also impacted on magazine approach behaviour as there was an effect of training ($F_{(2,42)}$ = 5.55, p<0.05) but no effect of intended devaluation or any

interaction (both F's<1). Tukey tests revealed that responding in the Up (p<0.05) and Down groups (p<0.05) was elevated compared to the Same group but the two groups that experienced a change in reward magnitude responded at equivalent rates (p=0.84) (mean magazine entries per minute (± SD): Same group = 8.0 (± 2.0); Up group = 11.6 (± 4.1); Down group = 10.95 (± 3.0)).

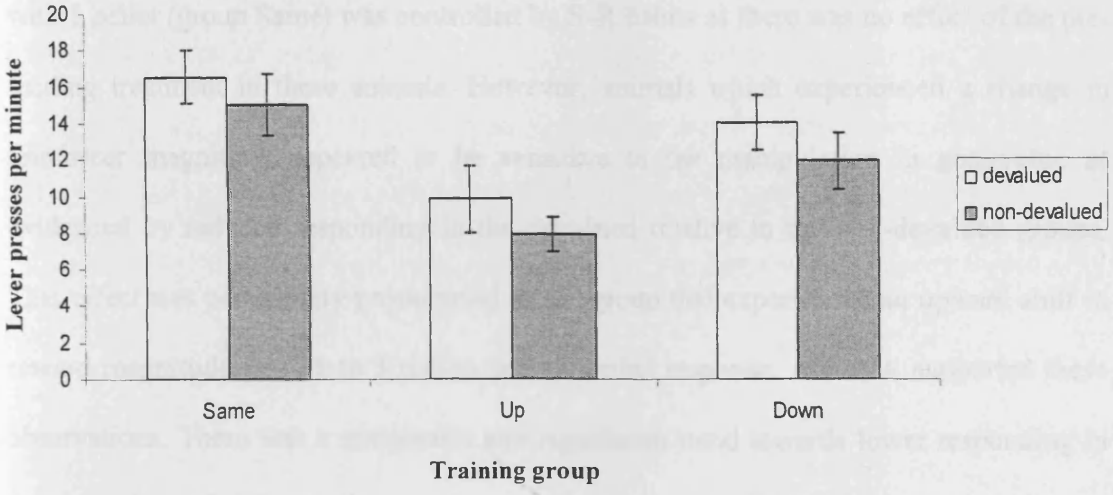

Figure 3.3.2 Mean lever presses per minute (±SEM) after no change (Same), an upward (Up) or downward shift (Down) in reinforcer magnitude.

A within subject comparison of lever press rates on the final day of acquisition and after the change in reinforcer magnitude (compare Figures 3.3.1 and 3.3.2) confirmed the change disrupted response rates ($F_{(1,42)}$ = 5.6, p<0.05). Significantly, there was a session x change interaction ($F_{(2,42)}$ = 19.8, p<0.05). Simple effects analysis of this interaction revealed that response rates in the Same group were comparable across the 2 sessions (F<1) but the increase in reward magnitude led to a marked reduction in rates in the Up

group relative to baseline ($F_{(1,14)}$ = 40.4, p<0.05). On the other hand, the reduction in reward magnitude resulted in higher rates in the Down group relative to baseline performance ($F_{(1,14)}$ = 4.73, p<0.05).

Extinction test – lever press performance. The mean lever presses per minute as a proportion of responding in session 10 test are displayed in Figure 3.3.3. Inspection of this figure suggests that the responding of animals which had been consistently reinforced with 1 pellet (group Same) was controlled by S-R habits as there was no effect of the pre-feeding treatment in these animals. However, animals which experienced a change in reinforcer magnitude appeared to be sensitive to the manipulation in goal-value as evidenced by reduced responding in the devalued relative to the non-devalued groups. This effect was particularly pronounced in the group that experienced an upward shift in reward magnitude from 1 to 3 pellets per rewarded response. ANOVA supported these observations. There was a marginally non-significant trend towards lower responding in the animals prefed the instrumental outcome ($F_{(1,42)}$ = 3.89, p=0.06) and a significant effect of the change in reinforcer magnitude ($F_{(2,42)}$ = 10.90, p<0.05). Tukey tests demonstrated that this effect reflected overall higher rates in the Up group compared to both the Same (p<0.05) and Down groups (p<0.05). Furthermore, ANOVA confirmed that the change in reward magnitude led to differential sensitivity of lever pressing to the reinforcer devaluation procedure ($F_{(2,42)}$ = 3.41, p<0.05). Simple effects analysis of this interaction supported the impression from Figure 3.3.3 that there was no effect of devaluation in the Same group (F<1). Although there was no statistical evidence for an effect in the Down group ($F_{(1,14)}$ = 1.83, p=0.2), the upward shift in reinforcer magnitude restored goal-sensitivity as there was a selective suppression in responding by animals

prefed the instrumental outcome compared to animals prefed the alternative reinforcer ($F_{(1,14)}$ = 8.21, p<0.05). Furthermore, there was no effect of the change in reinforcer magnitude on devalued performance ($F_{(2,21)}$ = 1.67, p=0.21) but non-devalued performance did differ across the 3 training groups ($F_{(2,21)}$ = 12.63, p<0.05). Subsequent analysis of this effect with Tukey tests revealed that non-devalued performance in the Up group was elevated compared to the Same (p<0.05) and Down groups (p<0.05). These differences may reflect non-selective drive effects of the pre-feeding manipulation on habitual performance compared to the selective effects seen on goal-directed responding in the Up group. Alternatively, the effect on non-devalued performance may be due to differences in baseline rates (see Figure 3.3.2). However the differential impact of the devaluation procedure on performance can not be attributed to baseline differences in response rates.

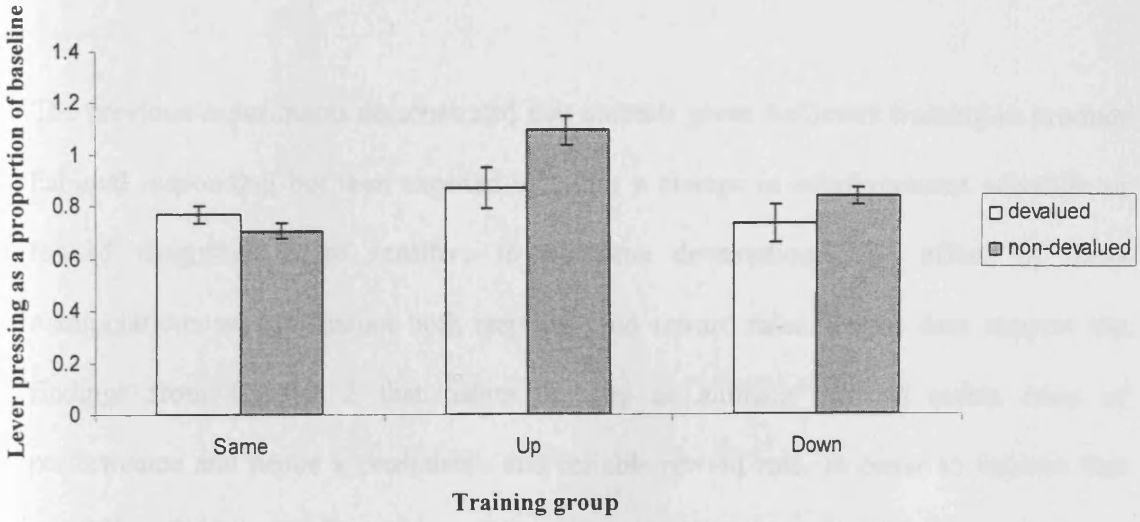

Figure 3.3.3 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after pre-feeding with the instrumental outcome (devalued - white bars) or an alternative reinforcer (non-devalued - grey bars).

Extinction test – magazine entry performance. Magazine entry behaviour was unaffected by any of the experimental factors. There was no effect of the change in reinforcer magnitude ($F<1$), devaluation ($F<1$) nor any interaction ($F<1$). (Mean magazine entries per minute as a proportion of baseline (± SD): Same devalued group = 0.64 (± 0.35); Same non-devalued group = 0.67 (± 0.29); Up devalued group = 0.81 (± 0.45); Up non-devalued group = 0.73 (± 0.36); Down devalued group = 0.54 (± 0.23); Down non-devalued group = 0.68 (± 0.35).

Experiment 6. The effect of a varying RI schedule on the sensitivity of an overtrained instrumental response to outcome devaluation.

The previous experiments demonstrated that animals given sufficient training to produce habitual responding but then exposed to either a change in reinforcement schedule or reward magnitude were sensitive to outcome devaluation. The effect of these manipulations was to disrupt both response and reward rates. These data support the findings from Chapter 2 that habits develop as animals achieve stable rates of performance and hence a predictable and reliable reward rate. In order to explore this possibility further, the Experiment 6 sought to prevent animals from experiencing a consistent reward rate by exposing them to a constantly varying schedule of reinforcement. If S-R habits develop as animals establish a predictable and reliable relationship between behaviour and reward, then animals that do not experience a consistent schedule of reinforcement should remain sensitive to changes in reward value even after extended training. Experiment 6 compared the sensitivity to outcome devaluation of animals that had received the same number of response-reward pairings (360) but were trained on either a static or varying reinforcement schedule.

Method and Materials

Subjects

The subjects were 48 experimentally naïve male Lister-hooded rats of mean weight 281g at the beginning of the experiment (range 244 – 312g). Animal care and handling was the same as in previous experiments.

Behavioural apparatus

The experiment was conducted in the same chambers as described for Experiment 3.

Behavioural Training

Magazine training. Magazine training proceeded as described previously.

Lever press training. As described elsewhere, animals received 2 days of CRF training. Thereafter half the animals (group Static) received identical training sessions to the High group in Experiment 1. Thus, the Static group received 9 sessions of training on a RI30, schedule earning 40 reinforcers in each session. The experimental group (group Varying) also received 9 sessions earning 40 reinforcers in each. However these animals were exposed to a reinforcement schedule that varied during each of these sessions. The computer control system randomly selected a value in a range from 5 to 120 (in increments of 5) and this value was set as the RI schedule for a two-minute period. After each two minute period had passed, the computer reset the RI schedule and a new value was randomly selected. This process continued until the animal had earned 40 reinforcers. Thus the maximum number of reinforcers an animal could potentially earn in any given two-minute period varied considerably (from 1 to 24) but the average of this range was approximately equivalent to the maximum number of reinforcers obtainable on a RI30 schedule in a two-minute period (4).

Devaluation by specific satiety

All animals then received 1 session of specific-satiety devaluation extinction testing during which lever press and magazine entry behaviours were assessed. Devaluation by specific satiety and extinction testing proceeded in the same manner as the other experiments in Chapter 3.

Data analysis

Statistical analysis was performed using analysis of variance (ANOVA) with between subject factors of reinforcement schedule employed during training (either static or varying) and devaluation by specific satiety (devalued versus non-devalued).

Results

Instrumental training. Training proceeded smoothly with all animals acquiring the instrumental response. By the end of the nine sessions, there was overall higher rates of responding in the Varying group compared to the Static group (mean lever presses per minute (±SD): Varying group = 20.3 (±7. 6); Static group = 15.7 (±6.0)). ANOVA revealed a main effect of training regime ($F_{(1,44)} = 5.44$, $p<0.05$) but no effect of intended devaluation ($F<1$) or an interaction between these factors ($F_{(1,44)} = 1.91$, $p=0.17$). Similarly on the final day of acquisition there were marginally higher rates of magazine entry in the Varying group (mean magazine entries per minute (±SD): Varying group = 12.2 (±4.7); Static group = 10.3(±5.1)) but the difference failed to reach statistical significance ($F_{(1,44)} = 1.77$, $p=0.19$). Furthermore, there was no effect of intended devaluation group on baseline magazine approach behaviour or an interaction (both

$Fs<1$). It is significant to note that as well responding at higher rates, the Varying group experienced greater variation in response rates across the 9 sessions of training compared to the Static group. This was confirmed by analysis of the standard deviation of response rates across the 9 sessions of instrumental training ($t_{(46)} = 2.6$ $p<0.05$) (mean variation in response rates (±SD) Varying group = 5.07 (±2.24); Static group = 3.66 (±1.45). Similarly, a t-test confirmed that the mean variation in the rate of reinforcement across the 9 sessions of training was statistically significantly higher in the Varying group compared to the Static group ($t_{(46)} = 7.35$, $p<0.05$) (mean variation in reinforcement rate (±SD) Varying group = 0.93 (±0.38); Static group = 0.34 (0.11)).

Extinction test – lever press performance. Figure 3.4.1 displays the lever press per minute as a proportion of baseline responding for both training groups during the 10-minute extinction test. The impression from this figure is that the devaluation treatment had a non-selective effect on responding in the Static group. Rates were comparable across the two devaluation groups and as such can be taken as evidence of response control by S-R habits. In stark contrast however, animals with the equivalent exposure to the reinforcer as the Static group but earned on a constantly varying RI schedule, appeared to be goal-directed as evidenced by a selective depression in responding by animals prefed the instrumental outcome. This description of the data was supported by ANOVA which yielded no effect of reinforcement schedule ($F_{(1,44)} = 1.67$, $p=0.2$) nor devaluation ($F_{(1,44)} = 3.82$, $p=0.06$) but critically an interaction between these factors ($F_{(1,44)} = 5.49$, $p<0.05$). Further analysis with simple effects confirmed a robust

devaluation effect in the Varying group ($F_{(1,22)} = 9.0$, $p<0.05$) but none in the Static group ($F<1$).

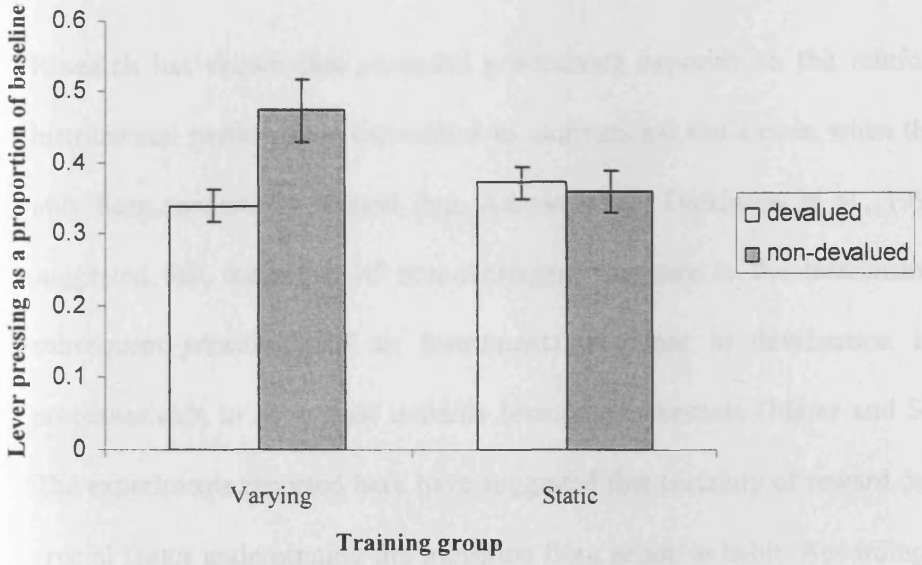

Figure 3.4.1 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after pre-feeding with the instrumental outcome (devalued - white bars) or an alternative reinforcer (non-devalued - grey bars).

Extinction test – magazine entry performance. Analysis of magazine approach behaviour during the extinction test revealed no effect of reinforcement schedule ($F<1$) but an effect of devaluation ($F_{(1,44)} = 6.16$, $p<0.05$). Significantly, this effect was unaffected by the reinforcement schedule employed during training as there was no interaction between these factors ($F<1$) (Mean magazine entries per minute as a proportion of baseline (±SD): Varying devalued group = 0.47 (±0.29); Varying non-devalued group = 0.86 (±0.74); Static devalued group =0.58 (±0.23); Static non-devalued group =0.89 (±0.68)).

Experiment 7. The effect of varying the rate of reward delivery on the disruption of goal-directed actions by pre-exposure to the reinforcer.

Research has shown that extended pre-training exposure to the reinforcer can render instrumental performance insensitive to motivational shifts even when the response has only been moderately trained (e.g. Adams, 1982; Dickinson et al., 1995). It has been suggested that the effect of non-contingent exposure to the instrumental outcome on subsequent sensitivity of an instrumental response to devaluation is the result of processes akin to those that underlie learned helplessness (Maier and Seligman, 1976). The experiments reported here have suggested that certainty of reward delivery may be a crucial factor underpinning the transition from action to habit. According to this position as the reward rate becomes well-predicted, response control is relinquished to the habit system. Thus, a possible alternative interpretation of the pre-exposure effect may be that non-contingent reinforcer presentations establish a context in which the rate of reward delivery is certain. Experiment 6 demonstrated that varying the reinforcement schedule during instrumental training disrupted the normal process whereby S-R habits come to dominate behavioural expression. The current experiment sought to examine whether manipulating the rate of non-contingent reinforcer presentations would prevent the accelerated formation of S-R habits after pre-training exposure to the instrumental outcome. Thus in Experiment 7, the sensitivity of instrumental responding to devaluation by specific satiety was indexed after extended pre-training exposure to the reinforcer delivered at a constant or varying rate. This was compared to the performance of animals that received no pre-training exposure.

Method and Materials

Subjects

The subjects were 48 experimentally naïve male Lister-hooded rats of mean weight 243g at the beginning of the experiment (range 213 – 289g). Animal care and handling was the same as in previous experiments.

Behavioural apparatus

The experiment was conducted in the same chambers as the ones described in Chapter 2.

Behavioural Training

Magazine training. Magazine training proceeded as described previously.

CRF training. After 2 days of magazine training, all animals received one session of lever press training reinforced on a CRF schedule. The session terminated when animals had earned a total of 50 reinforcers.

Extended magazine training. The control animals (no extended exposure group) received no training for the next 9 days and remained in their home cages. Animals in the Static and Varying groups received 9 days of non-contingent reinforcer presentations. Animals were placed in the conditioning chambers and the houselight came on to mark the beginning of the session. The lever was not available during these sessions. Animals then received 40 non-contingent reward deliveries either on a fixed RT-30 seconds schedule

(group Static) or a varying RT schedule (group Varying). As in the previous experiment, every 2 minutes the computer randomly selected a number between 5 and 120 and until the next 2-minute period had elapsed, this value represented the RT schedule (see experiment 6 for details). Once the animals had received 40 presentations, the houselight went out and the session was terminated.

Lever press training. After the final day of extended magazine training, all animals (including controls) received one session of instrumental training reinforced on a RR5 schedule. As with previous experiments, animals earned 40 reinforcers in this session.

Devaluation by specific satiety.

One day after the RR5 training, an extinction testing following devaluation by specific satiety was conducted.

Data analysis

Statistical analysis was performed using analysis of variance (ANOVA) with between subject factors of pre-training exposure (static rate of exposure, varying rate of exposure or no extended exposure) and devaluation by specific satiety (devalued versus non-devalued).

Results

Instrumental training. All animals acquired the instrumental response during the single RR5 session at the same rate. ANOVA yielded no effect extended exposure to the

reinforcer ($F_{(2,42)}$ = 1.13, p=0.33), intended devaluation ($F_{(1,42)}$ = 1.73, p=0.2) nor an interaction between these factors (F<1) (Mean lever presses per minute (±SD): Varying group = 10.4 (±3.6); Static group = 11.0 (±4.1); no exposure = 12.4 (±4.1). Similarly magazine entry behaviour was unaffected by any of these factors (all F's<1) (Mean magazine entries per minute (±SD): Varying group = 6.5 (±5.1); Static group = 5.9 (±3.0); no exposure = 6.9 (±2.9).

Extinction test – lever press performance. The lever presses per minute as a proportion of baseline are presented in Figure 3.4.1. It is clear from this figure that in contrast to previous reports (e.g. Adams, 1982) non-contingent exposure to the reinforcer failed to disrupt the selective sensitivity of less pressing to outcome devaluation. Similarly, the rate of delivery during reward exposure had no effect on lever press performance and if anything sensitivity to outcome devaluation was attenuated in the Varying group. ANOVA confirmed that instrumental performance in all animals was controlled by goal-directed A-O associations as there was an effect of devaluation ($F_{(1,42)}$ = 9.64, p<0.05) but no effect of pre-training exposure. Significantly, there was no evidence of differential sensitivity to outcome devaluation across the three training groups as there was no devaluation x exposure interaction (both Fs<1).

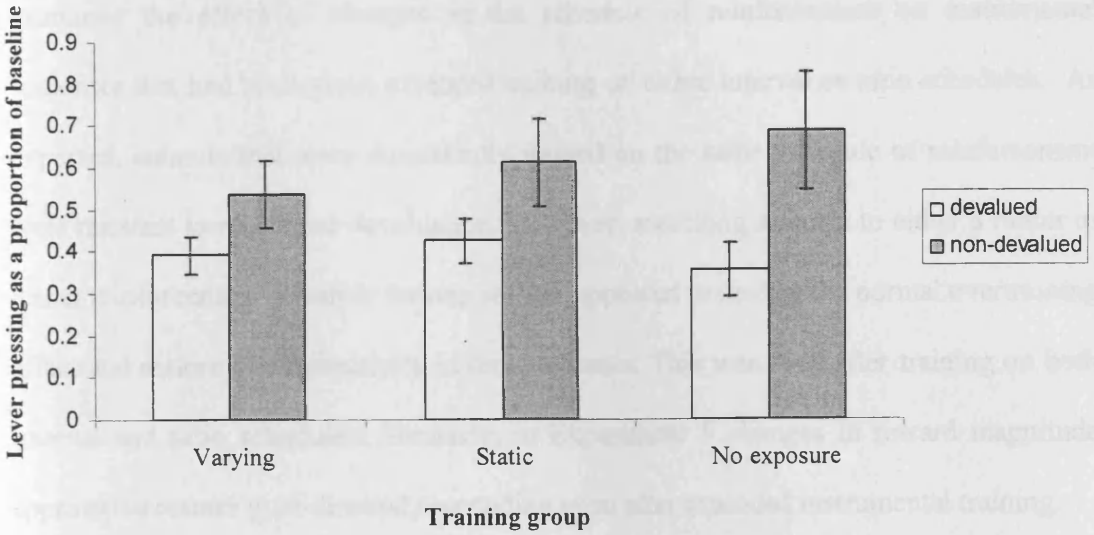

Figure 3.4.1 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after pre-feeding with the instrumental outcome (devalued - white bars) or an alternative reinforcer (non-devalued - grey bars).

Extinction test – magazine entry performance. Similarly, magazine approach behaviour during the extinction test was guided by outcome expectancy in all groups as evidenced by an effect of devaluation ($F_{(1,42)} = 5.96$, $p<0.05$) but no effect of pre-exposure ($F_{(2,42)} = 1.65$, $p=0.21$), nor any interaction ($F<1$).

Discussion

The experiments reported here explored the effect of various manipulations on the sensitivity of overtrained lever press performance to outcome devaluation either by specific satiety or conditioned taste aversion (Experiment 3b only). Experiments 3 and 4

examined the effect of changes in the schedule of reinforcement on instrumental responses that had been given extended training on either interval or ratio schedules. As expected, animals that were consistently trained on the same schedule of reinforcement were resistant to reinforcer devaluation. However, switching animals to either a richer or leaner reinforcement schedule for one session appeared to disrupt the normal overtraining effect and restore goal-sensitivity in these animals. This was seen after training on both interval and ratio schedules. Similarly, in Experiment 5 changes in reward magnitude appeared to restore goal-directed responding even after extended instrumental training.

It is striking that despite extensive training, animals in Experiments 3-5 were acutely sensitive to changes to the response-reward contingency. As such, these results suggest that overtraining an instrumental response does not necessarily degrade the response-reward relationship or that an animal's knowledge of this relationship changes over the course of training. However, following previous suggestions that habits develop as animals no longer experience the behaviour-reward correlation (e.g. Dickinson, 1985), it could of course be argued that the manipulations to the schedule of reinforcement in Experiments 3-4 provided animals with renewed experience of this relationship and hence account for these animals' sensitivity to reinforcer devaluation. For example, it is clear that the increase in the schedule of reinforcement in Experiments 3a and 4 (groups RI9 and RR3) re-exposed animals to a strong positive relationship between behaviour and reward. However, animals that experienced a decrease in the schedule of reinforcement (RI100 groups) were equally sensitive to goal-value at test, even though the shift to a leaner reinforcement schedule may have further weakened the relationship between

behaviour and reward. Furthermore, Experiment 5 did not change the behaviour-reward function *per se,* as responses continued to be rewarded on a RI30 schedule of reinforcement and yet these animals were sensitive to reinforcer devaluation. As these manipulations did not influence the behaviour-reward relationship uniformly (i.e. some manipulations may have strengthened this relationship but others may have weakened it), it is unlikely that their effects on this relationship can account for the restoration of goal-responding observed in Experiments 3-5.

The effect of all these manipulations was to disrupt well-established instrumental performance. Chapter 2 identified variation in response rates as critical to the development of S-R habits. In Experiments 1 and 2, habits appeared to develop as response rates ceased to vary. This would suggest that animals in the current experiments were sensitive to outcome value at test because manipulations to the schedule of reinforcement and reward magnitude disrupted the stable rates of responding that had been established over the previous 9 sessions of instrumental training. This position is further supported by the finding that decreasing reward magnitude in Experiment 5 (from 3 to 1) did not disrupt instrumental performance to the same extent as the upward shift and at test the Down group showed little or no sensitivity to outcome devaluation. Experiment 6 took a different approach; rather than disrupting well-established instrumental performance, it sought to offset the development of S-R habits by exposing animals to a constantly changing schedule of reinforcement during training. If the development of habits is simply the product of extended response practice or exposure to the reinforcer, then animals in the Varying group in Experiment 6 should have shown the

normal overtraining effect of resistance to reinforcer devaluation. However, instrumental performance in these animals remained sensitive to changes in goal-value even after 9 sessions of training. This manipulation prevented animals from ever establishing and experiencing stable response and reward rates. Indeed, animals in the Varying group experienced greater variability in response rates across the 9 sessions of instrumental training compared to the group consistently reinforced on a RI30 schedule of reinforcement. As such, the results from Experiment 6 provide further evidence that variation in these factors is critical to the transition from action to habit. A further notable feature of this experiment was the finding that the Varying group produced higher rates of responding during acquisition compared to controls. This is significant because it contrasts with the effects on response rates of the changes to reinforcement schedule and reward magnitude in Experiments 3-5, as these manipulations tended to lower rather than increase rates. Despite these higher rates, animals in Experiment 6 failed to develop habitual responding. This demonstrates that whether or not these manipulations increased or decreased baseline response rates, they produced equivalent effects on sensitivity to outcome devaluation at test.

The current experiments provide important insights into the behavioural processes involved in the development of S-R habits. They demonstrate unequivocally that habits are not simply the product of response repetition. For example, animals in the RI100 groups in Experiment 3a and 3b, as well as the Varying group in Experiment 6, pressed the lever significantly more than control animals and yet were still sensitive to outcome devaluation at test. On the other hand, these results also show that low response rates are

not necessarily associated with sensitivity to reinforcer devaluation. Thus neither the extent of response practice nor the absolute level of responding predicted animals' sensitivity to changes in reward value. Furthermore, these data demonstrate that increased exposure to the reinforcer is not sufficient to render instrumental performance habitual. In all the current experiments, animals received equivalent exposure to the reinforcer as controls (400 rewards in Experiments 3-5, 360 rewards in Experiment 6) but still showed differentially sensitivity to outcome devaluation at test. This is further supported by the results from the final experiment in this chapter. In contrast to previous research, Experiment 7 failed to find any evidence that extended pre-training exposure to the reinforcer facilitates habit formation.

What these experiments do demonstrate is that variation in response and reward rates are critical to the establishment of S-R habits. Manipulations that disrupted instrumental performance appeared to restore goal-directed responding in overtrained animals. As such, these findings are consistent with those from Chapter 2 that revealed that habitual responding emerged as both within and between session variation in response rates declined. These results are also consistent with the suggestion that habits develop as a stable rate of responding becomes associated with a reliable and well-predicted reward rate. Disrupting stable rates would render moribund the reliable estimation of the relationship between a certain level of responding and reward established over the previous 9 sessions of training. For example, the sudden switch in reinforcement schedule would presumably lead to the unexpected occurrence or omission of rewards. Similarly, the constantly varying reinforcement schedule in Experiment 6 would prevent animals

from establishing a stable and predictable relationship between their behaviour and reward delivery. These issues will be considered in greater depth in Chapter 6.

It is perhaps important to address the reasons why the manipulations introduced in Experiments 3-5 disrupted instrumental performance. Interestingly, the effect of the changes in reinforcement schedule introduced in Experiments 3-4 was to lower response rates relative to baseline regardless of whether the schedule of reinforcement was increased or decreased. Although the decline in response rates that accompanied the switch to a leaner rate of reinforcement is predicted by both the Law of Effect and molar correlational theories of instrumental performance (e.g. Baum, 1981), the finding that increasing the rate of reinforcement led to reduced response rates is perhaps surprising. However, there is some evidence that increasing the reinforcement schedule on both interval and ratio schedules does cause a downturn in response rates (e.g. Allison, 1980; Baum, 1981). It has been argued that the downturn in response rates as the reinforcement schedule is increased, may reflect an attempt to maintain a certain reinforcement rate (e.g. Ettinger at al., 1987) or more simply greater time spent consuming rewards (Baum, 1981). Whatever the merits of these arguments, for the current discussion it is significant that the effect of these manipulations was to disrupt response rates.

Similarly in Experiment 5, changes in reward magnitude appeared to disrupt instrumental performance. Previous investigations into the effect of varying reward magnitude on instrumental responding have produced mixed results, with some research showing that increasing reward magnitude tends to lead to higher response rates while others have

found the opposite pattern of results (see Reed, 1991). Nevertheless, the finding in Experiment 5 that animals reinforced with 3 pellets (Down group) during acquisition pressed the lever at rates lower than animals earning 1 pellet (Same and Up groups) is consistent with the results from Reed's study (1991) that on interval schedules, but not ratio schedules, larger reward magnitudes are associated with lower response rates. This was confirmed in Experiment 5 when animals experienced a change in reward magnitude on the 10^{th} day of instrumental training. In this session, animals that had previously earned 1 pellet per rewarded response, now received 3 and rates in these animals dropped compared to baseline levels. Conversely, when the number of reinforcers delivered per rewarded response was increased (Up group from 1 to 3 pellets per rewarded response), there was a decline in response rates relative to baseline. Thus the effect of the changes in reward magnitude in the 10^{th} session was either to increase or decrease response rates relative to those established over the 9 previous sessions. Significantly in a subsequent test of goal-sensitivity, responding in animals that experienced a shift in reward magnitude was shown to be guided by the current value of the reinforcer. Perhaps more revealing was the finding that animals in the Up group, which experienced the greatest change in response rates relative to baseline, showed the greatest sensitivity to goal-value at test.

In summary, the findings from the current experiments support the conclusion from Chapter 2 that variability in behaviour is a critical determinant of the transition from action to habit. Furthermore, these results provide behavioural evidence that A-O associations are not irretrievably lost over the course of instrumental training, and that A-

O processes can exert control of behaviour that would normally be modulated by S-R mechanisms. Thus at both the psychological and neural level, there is considerable evidence for the existence of systems that support both the performance of goal-directed actions and S-R habits. Significantly, the current findings highlight factors that contribute to the dominance of one system over the other.

CHAPTER 4

Experiments 8-9: The effect of pre- and post-training amphetamine exposure on the sensitivity of a moderately trained instrumental response to outcome devaluation.

As discussed in the Chapter 1, there is good evidence from human studies that dopamine is involved in the transition from goal-directed actions to habits. For example, the finding that patients with Parkinson's disease are impaired on procedural learning tasks would suggest that dopamine is fundamental to the consolidation and execution of S-R learning. However, evidence from studies involving laboratory rats has been far from definitive. The imperspicuity of the evidence obtained from rodents does not negate the findings from human studies but is more the product of methodological shortcomings, and in particular, the failure to probe the associative structure underpinning apparent habitual behaviours. The aim of the current experiments, therefore, was to seek further empirical support for dopaminergic modulation of habit formation. By testing animals in a reinforcer devaluation task, the associative structure of any observed behaviour could be explicitly demonstrated and hence the procedural failings of previous experiments could be obviated.

In the current experiments we assessed the effect of sensitization of dopaminergic systems on the control of goal-directed behaviour and sought evidence that increased

dopamine transmission would enhance habit formation. Sensitization to the effects of psychomotor stimulants, such as amphetamine, develops following repeated exposure (e.g. Robinson and Becker, 1986). The main behavioural manifestation of sensitization is a progressive and enduring enhancement of the psychomotor activating effects of psychostimulants as well as a potentiation of their reinforcing and incentive motivational properties (Stewart and Badiani, 1993; Robinson and Berridge, 1993). Coupled with persisting behavioural consequences, repeated exposure to psychostimulants induces long-term neural adaptions within brain areas that subserve learning and memory functions, including the prefrontal cortex, amygdala and hippocampus as well as the mesolimbic and mesostriatal dopamine systems. (Vanderschuren and Kalivas, 2000; Everitt and Wolf, 2002; Robinson and Kolb, 2004). Indeed, it has been widely reported that sensitization with psychostimulants leads to enhanced appetitive Pavlovian conditioned responding (Harmer and Phillips, 1998; Harmer and Phillips, 1999; Taylor and Horger, 1999; Taylor and Jentsch, 2001; Phillips, Harmer and Hitchcott, 2002). Similarly repeated intermittent administration of combined D_1 and D_2 agonists, such as amphetamine, has been shown to produce exaggerated activity and severe streotypies and consequently sensitization has been widely used to model repetitive behaviours associated with neuropsychiatric disorders such as obsessive-compulsive disorder and Tourette's Syndrome (Canales and Graybiel, 2000; Capper-Loup, Canales, Kadaba and Graybiel, 2002; Saka, Goodrich, Harlan, Madras and Graybiel, 2004). Furthermore, a recent study has reported disruption of the effects of outcome devaluation on Pavlovian-conditioned magazine approach after sensitization with cocaine (Schoenbaum and Setlow, 2005). Taken together these findings suggest that sensitization may provide a

valid method of probing the neuropharmacological processes involved in the transition from action to habit.

The current experiments examined the effects of amphetamine pre-treatment on animals' ability to produce goal-directed actions, determining whether sensitization of dopaminergic systems accelerates the dominance of S-R habits. The performance of amphetamine-exposed rats was compared with vehicle controls in a reinforcer devaluation task after limited training. Dopamine has been implicated in many cognitive processes (e.g. Robbins, 2005) and has a well-established role in motor control; as such it is important to dissociate any effects on associative learning from non-associative processes such as attention, motivation and sensorimotor function. In concert with this, to assess whether effects were mediated by learning or performance of the instrumental response, we measured sensitivity to outcome devaluation in animals exposed to amphetamine before (Experiment 8) or after training (Experiment 9).

Materials and Methods

Experiments 8 and 9: Pre- and post-training amphetamine sensitization and instrumental devaluation by specific satiety and lithium chloride–induced nausea.

Subjects

32, naïve, male, hooded Lister rats (Harlan UK Ltd., Bicester, Oxon, UK) were used in Experiment 8. At the beginning of the experiment their mean *ad libitum* weight was 277

g (range, 255-323 g). The subjects in Experiment 9 were 32, naïve, male, hooded Lister rats, with a mean *ad libitum* weight of 288g (range, 275-318g). Animal care and holding conditions were as described elsewhere.

Drugs

d-Amphetamine sulphate (Sigma Chemical Co., Poole, UK) was dissolved in sterile phosphate buffered saline (PBS). PBS was also used for control vehicle injections. Doses of *d*-amphetamine sulphate, 2mg/kg (sensitizing treatment) and 0.5 mg/kg (activity assay), were calculated as the salt.

Apparatus

The experiments were conducted in the chambers described previously. The reinforcers used were 20% w/v sucrose solution flavoured with grape Kool-Aid (0.05% w/v) and 20% w/v maltodextrin solution flavoured with cherry Kool-Aid (0.05% w/v) (Cybercandy Ltd., London, UK). Pilot studies indicated that in normal rats these reinforcers were well matched for motivational value but could be easily discriminated.

Sensitization.

Rats received intra-peritoneal (*i.p.*) injections of 2mg/kg *d*-amphetamine sulphate (amphetamine sensitized group) or the equivalent volume of vehicle PBS (control group), once per day for 7 consecutive days. Rats were returned to their home cages immediately after each injection. Over a seven-day injection-free period, animals in Experiment 8 were reduced to 80% of their *ad libitum* weight, prior to the start of behavioural training. One rat in Experiment 8 died during sensitization treatment so that 31 rats in total (15 vehicle controls and 16 amphetamine exposed rats) proceeded to the training stage.

Animals in Experiment 9 were reduced to 80% of their *ad libitum* weight before undergoing behavioural training. On the completion of this training, the rats received the sensitization treatment followed by a seven-day-injection-free period prior to testing. Although there was a minor difference in the period of time between cessation of amphetamine injections and the start of devaluation testing between Experiments 8 and 9 (14 and 11 days respectively), this is unlikely to influence the assessment of sensitization, which has been shown to have profound behavioural effects across weeks and even months (Vanderschuren and Kalivas, 2000).

Behavioural training

In Experiment 8, following the sensitization procedure each animal was assigned to one of the eight conditioning chambers, and thereafter was always trained in that chamber. At the start of each session, the house light came on and remained on throughout the session. The house light went out at the end of each session. Training consisted of two stages: magazine training and lever press training. This was followed by extinction tests after devaluation by specific satiety and lithium chloride-induced nausea. In Experiment 9, animals received training prior to the sensitization treatment, then received extinction tests after devaluation by lithium chloride-induced nausea.

Magazine training. Magazine training proceeded as described earlier. Half the animals were trained to collect the sucrose solution and the other half the maltodextrin solution (counter-balanced across treatment and devaluation groups).

Lever press training. Lever press training was conducted as for the Low group in Experiment 1. Thus after 2 sessions of CRF training, animals received 3 sessions on a RI30 schedule, earning 40 rewards in each session.

Alternative reinforcer. As we planned to test the animals' sensitivity to outcome devaluation, and in order to ensure that the non-devalued group readily consumed an alternative reinforcer in pre-feeding sessions, we equated the animals' exposure to two reinforcers. In addition to the reinforcers earned in lever press sessions, each rat received three sessions during which 40 presentations of the alternative reinforcer were made on a random time (RT) 30-second schedule. The lever was withdrawn during these sessions. Half the animals were exposed to the alternative reinforcer in the afternoon following morning lever press training, and the other half received alternative reinforcer sessions in the morning prior to lever press training in the afternoon (counter-balanced across treatment and devaluation groups).

Devaluation by specific satiety (Experiment 8 only)

All animals then received one session of devaluation by specific satiety followed by an extinction test during which lever presses and magazine entry behaviour were assessed. Animals were placed in feeding cages and given free access for one hour to either the instrumental outcome (Devalued group) or the alternative reinforcer (Non-devalued group). Immediately following this pre-feeding session, the animals were transferred to the conditioning chambers and received an 8-minute extinction test in the absence of

reward delivery. The lever was present during this session, but no reinforcers were delivered.

Devaluation by lithium chloride (Experiments 8 and 9)

Reminder session. As the animals had been through the extinction test above, rats in Experiment 8 received a reminder session on the day after the test. Animals were given one session in which they lever-pressed to earn a total of 40 rewards. The reminder session was identical to the initial sessions of instrumental training.

Thereafter, animals received three days of devaluation with lithium chloride (LiCl). On each day the rats were placed in the operant chambers and were given 40 free presentations of either the instrumental outcome (devalued group) or the alternative reinforcer (non-devalued group) on an RT 30-second schedule. Immediately after the cessation of each session, the devalued group received a 0.15M, 10 ml/kg *i.p.* injection of LiCl solution (Sigma Chemical Co., Poole, UK) and the non-devalued group an injection of the equivalent volume of saline. 24 hours after the final session of taste aversion training, animals' sensitivity to outcome devaluation was assessed in a further 8-minute extinction test in the absence of reward delivery. This was conducted as described above.

Re-acquisition test. In order to demonstrate that the devalued group had acquired an aversion to the instrumental outcome, all rats underwent a 15-minute re-acquisition test. The animals were placed in the conditioning chambers and lever pressed to earn the instrumental outcome on an RI 30-second schedule.

Activity assay. To confirm sensitization, all animals were administered a 0.5 mg/kg *i.p.* amphetamine challenge before assessment of levels of locomotor activity. These tests occurred immediately following the re-acquisition tests. Activity was monitored using eight chambers (56 cm wide x 39 cm deep x 19 cm high). Activity within each chamber was recorded with pairs of photobeams situated 20 cm apart and 18 cm from the end of the cage, connected to a control box (Paul Fray, Cambridge, UK). Each beam break resulted in an incremental count for that chamber and was recorded by an Acorn computer programmed in BBC Basic. Locomotor activity was measured (total number of photobeam breaks) for 30 minutes.

Data analysis

Statistical analysis was performed using analysis of variance (ANOVA) with between subject factors of devaluation (devalued versus non-devalued) and sensitization treatment (either sensitized or vehicle controls). As the standard deviation was proportional to the mean, the extinction data were subject to logarithmic transformation (Howell, 2002).

Results

Experiment 8: The effect of pre-training amphetamine exposure on the sensitivity of a moderately trained instrumental response to outcome devaluation by specific satiety.

Instrumental training. All the rats acquired the initial instrumental response at the same rate (data not shown). Significantly, by the end of the three days of training, there were no differences in baseline responding due to pre-treatment with amphetamine (mean responses per minute (±SD) vehicle to-be-devalued group = 10.3 (±3); vehicle to-be-valued group = 10.7 (±3.2); amphetamine to-be-devalued group = 9.2 (±2.2); amphetamine to-be-valued group = 9.2 (±2.8)). This was confirmed by ANOVA which revealed no effect of amphetamine treatment ($F<1$), devaluation group ($F_{(1,27)} =1.57$, $p=0.22$) nor an interaction between these two factors ($F<1$) and suggests that baseline differences are unlikely to account for any effects of amphetamine in subsequent extinction tests. Similarly, there was no effect of amphetamine exposure on magazine entry behaviour. There was no effect of treatment, intended devaluation or an interaction (highest $F_{(1,27)} = 2.807$, $p=0.11$) (mean magazine entries per minute (±SD) vehicle to-be-devalued group = 6.0 (±2.7); vehicle to-be-valued group = 5.5 (±1.5); amphetamine to-be-devalued group = 7.2 (±2.1); amphetamine to-be-valued group = 5.4 (±1.3)).

Lever press extinction test performance. The mean response rates per minute as a proportion of baseline (which did not differ – see above for details) for the eight minutes of the extinction test are presented in the Figure 4.1.1. This suggests that vehicle pre-treated animals' lever press performance was sensitive to the current value of the goal. Thus, the vehicle-injected control group performed fewer lever presses as a proportion of their baseline rates after pre-feeding with the instrumental outcome (devalued – white bars) compared to those pre-fed the alternative reinforcer (non-devalued – grey bars). Conversely, the performance of the amphetamine-sensitized animals was not goal-

directed as demonstrated by their failure to show sensitivity to the change in reward value. The devalued group pressed the lever at an equivalent rate to the non-devalued group, suggesting that their responding was insensitive to goal value and habitual.

This description of the data was confirmed by statistical analysis. ANOVA yielded no effect of devaluation ($F_{(1,27)}$=2.99, p=0.1) nor treatment (F<1), but critically a significant treatment x devaluation interaction ($F_{(1,27)}$ =4.23, p<0.05). Simple effects analysis revealed that devalued and non-devalued performance differed in the vehicle-injected control group ($F_{(1,13)}$=6.92, p<0.05) but not in the amphetamine-sensitized animals (F< 1).

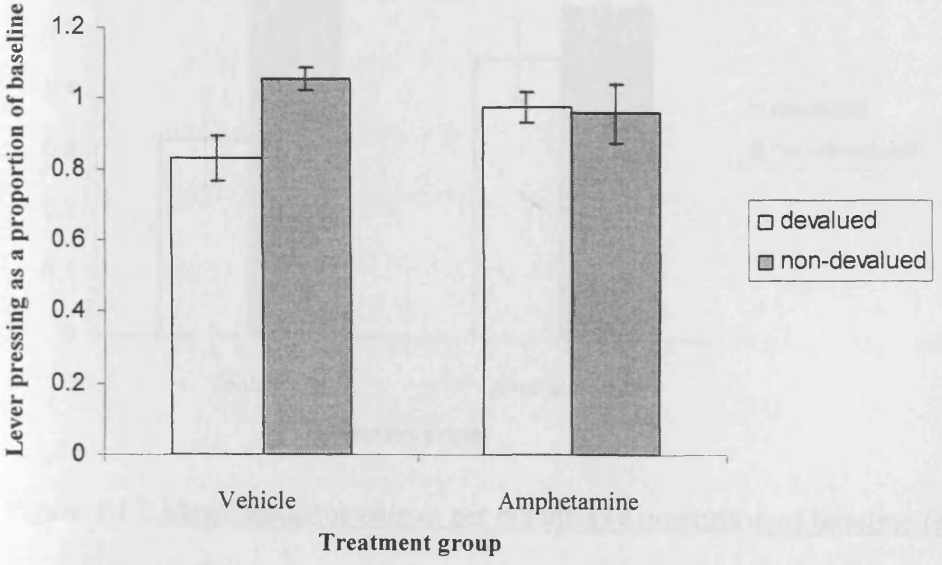

Figure 4.1.1 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after pre-feeding with the instrumental outcome (devalued - white bars) or an alternative reinforcer (non-devalued - grey bars).

Magazine entry extinction test performance. Figure 4.1.2 shows magazine entry behaviour during the extinction test. Preliminary analysis revealed no effect of treatment on baseline levels of magazine entry. This figure indicates that pre-feeding produced a decrease in magazine entry behaviour in both treatment groups but that this effect was more marked in the vehicle controls. Statistically, ANOVA revealed only a main effect of devaluation ($F_{(1,27)}$=4.96, p<0.05) but no effect of treatment (F<1) nor an interaction ($F_{(1,27)}$= 1.1, p=0.3).

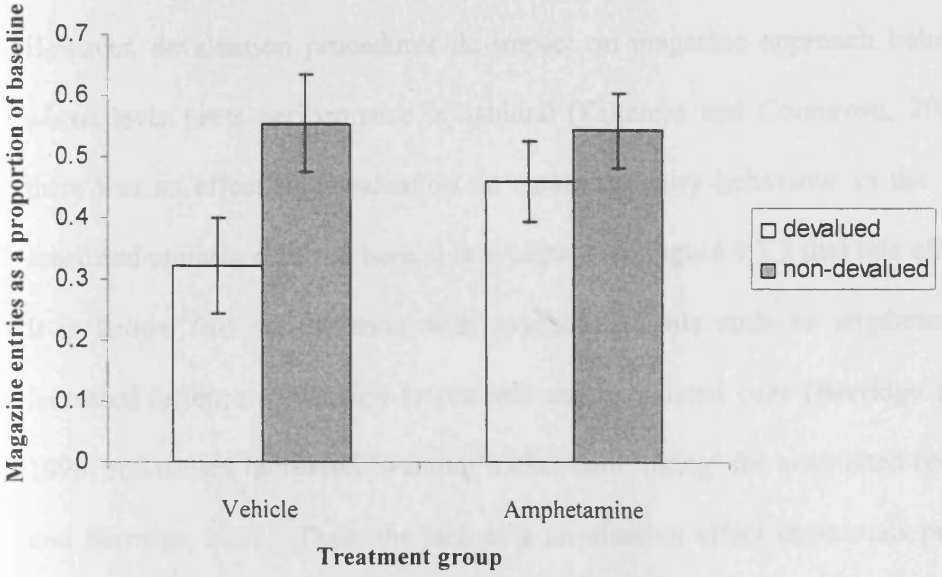

Figure 4.1.2 Mean magazine entries per minute as a proportion of baseline (±SEM) in the extinction test after pre-feeding with the instrumental outcome (devalued - white bars) or an alternative reinforcer (non-devalued - grey bars).

Experiment 8: The effect of pre-training amphetamine exposure on the sensitivity of a moderately trained instrumental response to outcome devaluation by lithium chloride.

In the test above, amphetamine-sensitized animals pre-fed the instrumental outcome showed no devaluation effect, indicating that responding was habitual rather than goal-directed. In this respect, the amphetamine sensitized rats' lever press performance mirrors that of overtrained rats (e.g. Adams, 1982) or the responding of animals with lesions to the prelimbic cortex (Balleine and Dickinson, 1998; Killcross and Coutureau, 2003). However, devaluation procedures do impact on magazine approach behaviour in rats' whose lever press performance is habitual (Killcross and Coutureau, 2003). Although there was an effect of devaluation on magazine entry behaviour in the amphetamine-sensitized animals reported here, it is evident from Figure 4.1.2 that this effect was small. It is known that sensitization with psychostimulants such as amphetamine leads to increased salience attribution to rewards and associated cues (Berridge and Robinson, 1998) and causes increased 'wanting' rather than 'liking' for associated rewards (Wyvell and Berridge, 2001). Thus, the lack of a devaluation effect in animals pre-treated with amphetamine shown above may not have been the result of the accelerated learning of S-R associations, but rather due to the failure of the pre-feeding procedure to devalue the outcome sufficiently. Conditioned taste aversion, induced by lithium chloride, produces far more robust devaluation effects in normal animals compared to pre-feeding with the instrumental outcome. In this test we re-assessed the animals' sensitivity to changes in

goal value but after pairing the reward with LiCl-induced illness. Furthermore, the level of aversion to the reinforcer could be assessed in a subsequent re-acquisition test.

Reminder session. As with the initial instrumental training, amphetamine-sensitized rats pressed at comparable rates to the vehicle controls (mean responses per minute (±SD) vehicle to-be-devalued group = 10.8 (±2.8); vehicle to-be-valued group = 7.9 (±2.5); amphetamine to-be-devalued group = 9.3 (±4.1); amphetamine to-be-valued group = 10.1 (±2.7)). Statistical analysis revealed no effect of treatment, devaluation or interaction (highest $F_{(1,27)} = 2.78, p = 0.11$).

Lever press extinction test performance. Figure 4.1.3 displays the instrumental performance during the 8-minute extinction test for the vehicle-injected control group and the amphetamine-sensitized rats as a proportion of their baseline responding. In the vehicle-injected control groups, test performance showed a marked reduction in responding after conditioned aversion training (devalued – white bars) relative to animals that had not been averted from that outcome (non-devalued – grey bars). By contrast, lever press performance of the amphetamine pre-treated group appeared to be impervious to the change in the goal value. The amphetamine-sensitized rats averted from the instrumental reward, showed comparable levels of responding to that of sensitized rats not averted from the reinforcer.

An ANOVA with treatment and devaluation as factors supported this observation. There was a main effect of treatment ($F_{(1,27)}= 9.71, p<0.05$) and of devaluation ($F_{(1,27)}= 11.63, p<0.05$) but crucially also a highly significant interaction between these two factors

($F_{(1,27)}$ = 8.22, p<0.05). Simple effects analysis of this interaction confirmed that the devalued vehicle-injected group showed a marked suppression in lever press responding compared to the non-devalued vehicle-injected animals ($F_{(1,13)}$ = 19.05, p<0.05), but that there was no effect of devaluation in amphetamine-sensitized rats (F<1). Further simple effects analysis revealed an effect of amphetamine treatment in the devalued groups ($F_{(1,13)}$ = 17.31, p<0.05) but not in the non-devalued groups (F<1).

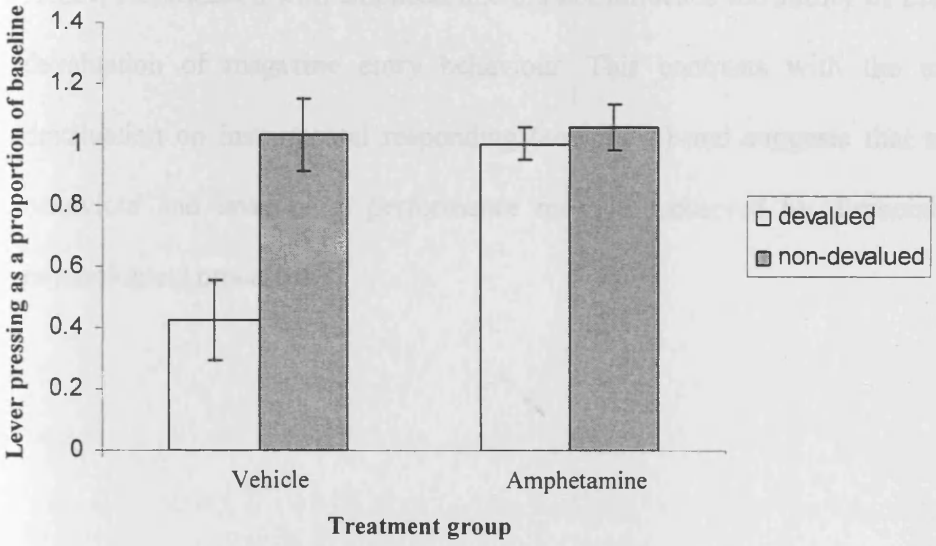

Figure 4.1.3 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after devaluation by LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Magazine entry extinction test performance The mean magazine entries per minute, as a proportion of baseline, during the extinction test following taste aversion training are shown in of Figure 4.1.4. Inspection of this figure reveals that the animals with an aversion to the reinforcer performed considerably fewer magazine entries compared to

the non-devalued controls. This was confirmed by ANOVA which yielded a highly significant main effect of devaluation ($F_{(1,27)} = 41.57$, $p<0.05$). It is also clear from Figure 4.1.4 that there were overall higher levels of magazine entry behaviour in the amphetamine-sensitized rats compared to the vehicle controls: ANOVA revealed a main effect of treatment ($F_{(1,27)} = 11.53$, $p<0.05$). However, the extent of the devaluation effect in the sensitized animals was equivalent to that seen in the vehicle-injected control animals, as demonstrated by the lack of a treatment x devaluation interaction ($F<1$). Hence, sensitization with amphetamine did not influence the ability of LiCl to produce a devaluation of magazine entry behaviour. This contrasts with the effects of LiCl devaluation on instrumental responding (see above) and suggests that magazine entry behaviour and lever press performance may be subserved by dissociable neural and psychological processes.

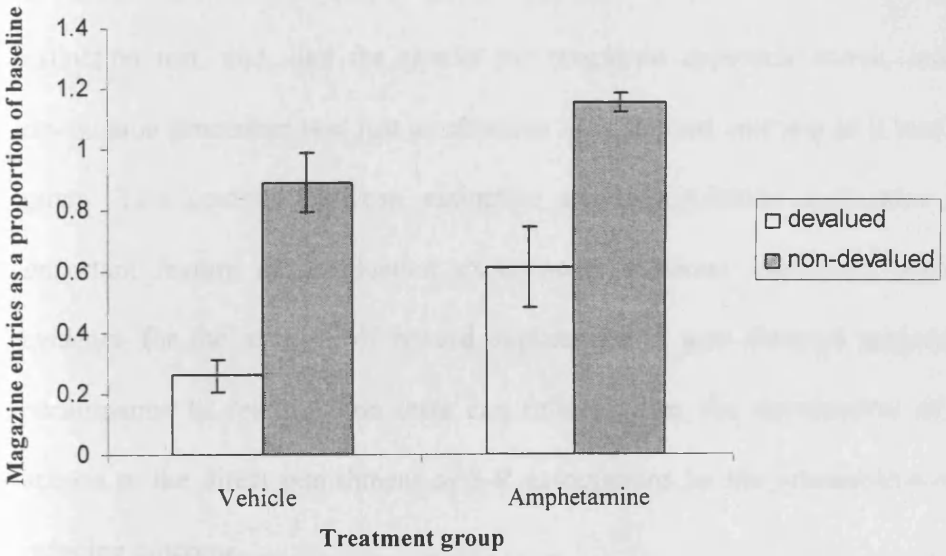

Figure 4.1.4 Mean magazine entries per minute as a proportion of baseline in the extinction test after devaluation by LiCl (devalued - white bars) or no devaluation (non-devalued grey bars).

Reacquisition test - Lever press performance. The results of the reacquisition test confirmed that the LiCl injections had successfully devalued the instrumental outcome in both drug treatment groups. The mean lever presses per minute for the rewarded re-acquisition test are presented in the Figure 4.1.5. This indicates that compared to the non-devalued control group, the devalued group performed considerably fewer lever presses in the 15-minute rewarded test. Statistical analysis by ANOVA produced a highly significant main effect of devaluation ($F_{(1,27)} = 89.75$, $p<0.05$). The trend towards higher levels of responding in the amphetamine-sensitized group was maintained in the reacquisition test ($F_{(1,27)} = 14.70$, $p<0.05$) but again the level of devaluation in these animals was comparable to that seen in the drug-naïve rats as there was no treatment x

devaluation interaction ($F<1$). These results stand in marked contrast to the results of the extinction test, and, like the results for magazine approach above, indicate that the devaluation procedure was just as effective in sensitized animals as it was in the control group. This contrast between extinction and reacquisition tests also highlights an important feature of devaluation experiments; whereas extinction tests can provide evidence for the strength of reward expectation in goal-directed responding, lowered performance in reacquisition tests can reflect either the devaluation of goal-directed actions or the direct punishment of S-R associations by the presentation of the nausea-inducing outcome.

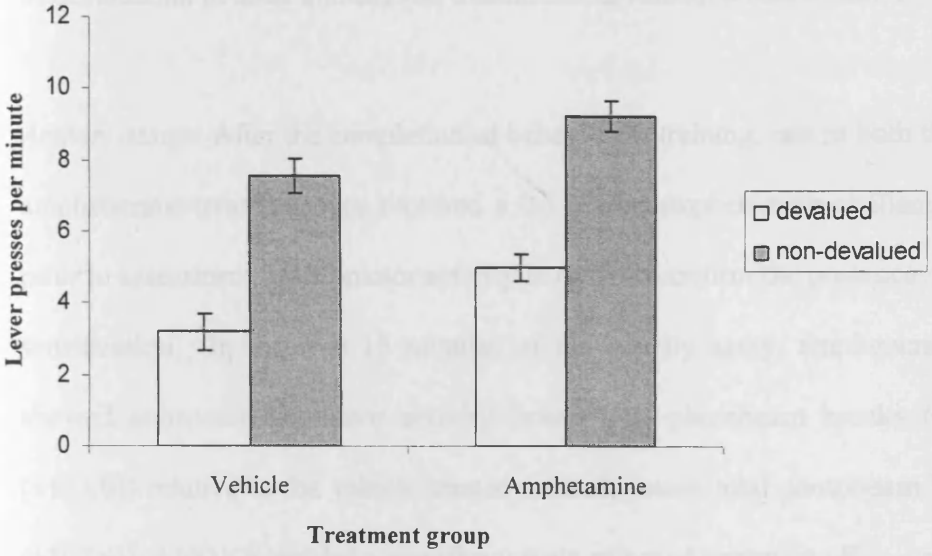

Figure 4.1.5 Mean lever presses per minute (±SEM) in the rewarded reacquisition test after devaluation with LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Reacquisition test – Magazine entry behaviour. The effectiveness of the LiCl treatment in devaluing the instrumental reward is further supported by analysis of magazine entry behaviour in the reacquisition test. The mean responses per minute (±SD) were: devalued vehicle group = 0.5 (±0.21); non-devalued vehicle group = 5.6 (±1.5); devalued amphetamine group = 2.8 (±1.1); non-devalued amphetamine group = 7.9(±4.5). Both devalued groups showed a marked suppression in magazine entries compared to the non-devalued groups. ANOVA revealed a main effect of devaluation ($F_{(1,27)} = 35.12$, $p<0.05$). The amphetamine-sensitized animals again displayed higher levels of magazine entry behaviour ($F_{(1,27)} = 6.88$, $p<0.05$), but this heightened activity did not influence the level of devaluation in these animals (no treatment x devaluation interaction, $F<1$).

Activity assay. After the completion of behavioural training, rats in both the vehicle- and amphetamine-treated groups received a 0.5 mg/kg amphetamine challenge immediately prior to assessment of locomotor activity in order to confirm the presence of psychomotor sensitization. In the first 15 minutes of the activity assay, amphetamine-treated rats showed enhanced locomotor activity (mean total photobeam breaks (±SD) = 314.5 (±111.9)) relative to the vehicle treated controls (mean total photobeam breaks = 230.1 (±107.4)). ANOVA yielded a significant main effect of treatment ($F_{(1,29)} = 4.59$, $p<0.05$); confirming that amphetamine pre-treatment had successfully sensitized these animals. We also examined correlations between the locomotor activity in response to the amphetamine challenge in the devalued sensitized animals and lever press performance in the 2 extinction tests. Locomotor activity in the sensitization assay bore no relationship

either to performance after devaluation by specific satiety (r= 0.1, p=0.81), or after devaluation by LiCl (r= 0.03, p=0.94). Hence, the failure to detect sensitivity to outcome devaluation in these animals is unlikely to be explained simply in terms of increased locomotor activity.

Experiment 9: The effect of post-training amphetamine exposure on the sensitivity of a moderately trained instrumental response to outcome devaluation by lithium chloride.

In Experiment 8 we demonstrated that animals that had been sensitized with amphetamine failed to alter their lever press performance in response to a change in the value of the reinforcer, brought about by pre-feeding with the instrumental outcome and pairing the reward with illness. These results indicate that in sensitized animals, instrumental responding was not goal-directed but rather stimulus-driven and habitual. However, whether the effect of sensitization was one on learning or performance of the response was confounded in this experiment. Sensitizing the animals after the initial training would allow these two possibilities to be dissociated. In Experiment 9, therefore, the sensitization treatment was conducted following initial lever press training. Following a week of recovery, the sensitivity of rats to outcome devaluation was assessed by pairing the reward with lithium chloride prior to an extinction test. If animals' responding in this test is independent of the current value of the goal, it would suggest that the sensitization treatment had had an effect on the performance or expression of the instrumental action. If the effect of sensitization is restricted to the acquisition phase of instrumental learning,

we would expect the animals to continue to show sensitivity to the changed value of the reinforcer.

Instrumental training. Training proceeded smoothly with all animals acquiring the instrumental response at the same rate (data not shown). By the end of training there were no differences in lever press responding between animals allocated to the amphetamine sensitization group and vehicle-injected control group, no effect of devaluation nor an interaction (highest $F_{(1,28)} = 2.19$, p=0.15) (mean responses per minute (±SD) vehicle to-be-devalued group = 11.0 (±1.6); vehicle to-be-valued group = 12.8 (±1.3); amphetamine to-be-devalued group = 12.0 (±3.3); amphetamine to-be-valued group = 11.7 (±1.1)).

Lever press extinction test performance. The mean lever presses per minute as a proportion of baseline for the 8 minutes of the extinction test are displayed in Figure 4.2.1. This suggests that irrespective of drug treatment, all animals with the devalued reinforcer (white bars) showed a marked reduction in lever pressing relative to the non-devalued group (grey bars). This description of the data was confirmed statistically by ANOVA with between subjects factors of sensitization treatment and devaluation group. The post-training amphetamine treatment had no effect on animals' sensitivity to outcome devaluation as there was a highly significant main effect of devaluation ($F_{(1,28)} = 43.1$, $p<0.05$) but no interaction ($F<1$). There was a trend for overall higher responding in the amphetamine sensitized group but it failed to reach the level of rejection of the null hypothesis ($F_{(1,28)} = 2.89$, $p=0.1$).

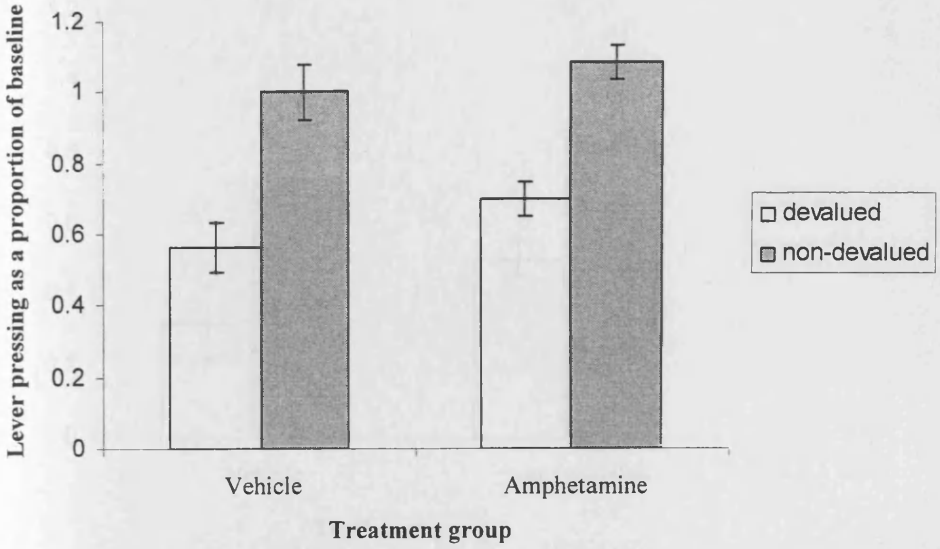

Figure 4.2.1 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after devaluation by LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Magazine entry extinction test performance. The success of the LiCl treatment in devaluing the outcome for both groups is further highlighted by analysis of magazine entry behaviour during the extinction test. As is clear from Figure 4.2.2, both the vehicle-injected control group showed a clear suppression in magazine activity compared to the non-devalued groups. ANOVA yielded a highly significant main effect of devaluation ($F_{(1,28)}$=21.45, p<0.05). Although there was a marginally-significant trend towards higher magazine activity in the amphetamine sensitized animals ($F_{(1,28)}$=4.17, p=0.051), this failed to impact on the level of devaluation in these animals as there was no treatment x devaluation interaction (F<1).

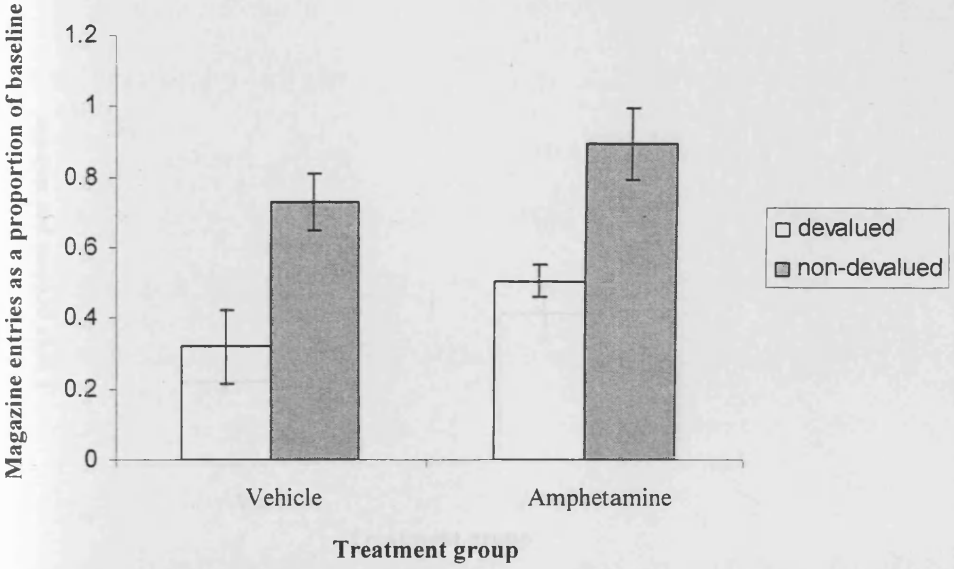

Figure 4.2.2 Mean magazine entries per minute as a proportion of baseline (±SEM) in the extinction test after devaluation by LiCl (devalued - white bars) or no devaluation (non-devalued grey bars).

Reacquisition test - Lever press performance. The results of the reacquisition test, displayed in Figure 4.2.3, confirmed that both treatment groups in the devalued condition had acquired a strong aversion to the reinforcer. Relative to the non-devalued controls, devalued rats pressed the lever at a lower rate. This observation was supported by statistical analysis. ANOVA revealed a main effect of devaluation ($F_{(1,28)} = 66.4$, $p<0.05$), and also a main effect of drug ($F_{(1,28)} = 17.12$, $p<0.05$) reflecting overall higher response rates in the amphetamine-sensitized animals. However, the higher level of responding in the amphetamine-sensitized rats did not influence the magnitude of the devaluation effect in these animals relative to vehicle controls, as there was no treatment x devaluation interaction ($F<1$).

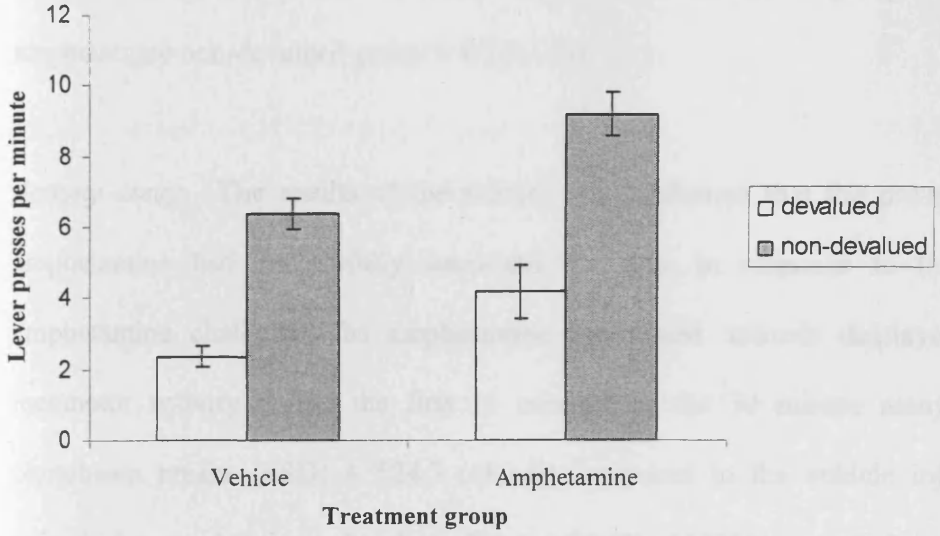

Figure 4.2.3 Mean lever presses per minute (±SEM) in the rewarded reacquisition test after devaluation with LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Reacquisition test – Magazine entry performance. Similarly, analysis of magazine approach behaviour during the reacquisition test confirmed that all animals irrespective of drug treatment had acauired a robust aversion to the reinforcer. ANOVA revealed an effect of devaluation ($F_{(1,28)}$ = 20.01, p<0.05) but also an effect of drug treatment reflecting overall higher levels of magazine entry activity in the amphetamine-sensitized animals ($F_{(1,28)}$ = 6.05, p<0.05). However, this did not impact on the level of devaluation in these animals as there was no treatment x devaluation interaction ($F_{(1,28)}$ = 1.1, p=0.3) (mean magazine entries per minute (±SD) vehicle devalued group = 0.71 (±0.76); vehicle

non-devalued group = 2.7 (±2.2); amphetamine devalued group = 1.5 (±1.5); amphetamine non-devalued group = 4.7 (±1.7)).

Activity assay. The results of the activity test confirmed that the pre-treatment with amphetamine had successfully sensitized the rats. In response to the 0.5 mg/kg amphetamine challenge, the amphetamine pre-treated animals displayed heightened locomotor activity during the first 15 minutes of the 30 minute assay (mean total photobeam breaks (±SD) = 524.7 (±80.4)) compared to the vehicle injected control animals (mean photobeam breaks = 386.6 (±88.2)). ANOVA revealed a main effect of treatment ($F_{(1,28)}$ = 21.43, p<0.05). This enhanced activity, moreover, did not correlate with test lever press performance (r= 0.02, p= 0.97).

Discussion

These experiments investigated the effects of amphetamine pre-treatment on the sensitivity of lever pressing to reward devaluation after limited training. Consistent with previous accounts, vehicle-injected control animals showed a selective suppression in lever press performance after reinforcer devaluation by either specific satiety or LiCl-induced nausea. However, pre-training exposure to amphetamine disrupted acquisition of goal-directed behaviour. Sensitized rats failed to modify their lever press performance in response to the changed value of the outcome, responding at equivalent levels to those

seen in non-devalued controls. This effect was observed after devaluation by both specific satiety and LiCl-induced nausea, and suggests that the control of responding in the amphetamine-treated rats was not dependent on the expected outcome but instead was dominated by reflexive habits. The studies also revealed a dissociation between the effect of pre- and post-training sensitization on sensitivity to outcome devaluation. Experiment 9 demonstrated that animals exposed to amphetamine after initial lever press training retained robust sensitivity to changes in reward value, indicating that amphetamine treatment disrupts the acquisition, but not expression, of goal-directed actions.

Several aspects of the current data deserve comment. The failure to detect sensitivity to outcome devaluation after amphetamine exposure in Experiment 8 cannot be accounted for in terms of a general learning impairment. Amphetamine-sensitized animals acquired the instrumental response at equivalent rates to the vehicle-injected control group. Moreover, these animals showed extinction at rates comparable to those seen in the non-devalued, vehicle pre-treated animals. There is evidence that antagonism of dopaminergic systems by neuroleptics produces response patterns that resemble extinction (Phillip and Fibiger; 1979, Gray and Wise, 1980), while amphetamine (Flecther, 1995, 1996; Foltin, 2004) and the D_2 agonist quinpirole (Kurylo and Tanguay, 2003) have been shown to attenuate extinction. However, in neither Experiment 8 nor 9 did amphetamine sensitization have any discernible effect on the rate of extinction with the experimental parameters we employed. Hence it is unlikely that the failure to detect sensitivity to outcome devaluation in Experiment 8 can be explained in terms of alterations in extinction processes at test by sensitization. Indeed, if this were the case we would have

expected to see equivalent results in Experiments 8 and 9, in contrast to the actual findings.

The locomotor activating effects of psychostimulants are also well documented (e.g. Stewart and Badiani, 1993). This was confirmed in the current experiments by heightened locomotor activity in response to an amphetamine challenge compared to drug-naïve performance. However, this increase in locomotor activity did not correlate with performance in the devaluation extinction tests. Similarly, the observation that magazine entry behaviour remained sensitive to changes in reward value, as well as the devaluation effect seen in the reacquisition test, confirm that the performance of amphetamine-sensitized animals was not simply a consequence of hyperactivity – they were able to suppress specific response tendencies in certain situations, and were not impaired in their general ability to inhibit responding. Thus, the results cannot be accounted for in terms of enhanced locomotor activity or general response perseveration.

Nor did amphetamine sensitization change the motivational and incentive impact of the devaluation treatments employed (see Wyvell and Berridge, 2001). It is clear from the magazine entry data and the reacquisition test that insensitivity to outcome devaluation was not due to any ineffectiveness of the pre-feeding treatment or a failure to acquire an aversion to the reinforcer after taste aversion training. In Experiment 8, while lever press performance in extinction was impervious to the shift in the value of the reinforcer, magazine entry behaviour remained sensitive to manipulations in goal value, and the reacquisition test confirmed that all animals had acquired an aversion to the reinforcer.

Similarly, as any explanation in terms of changes in the effectiveness of reward devaluation depends on effects restricted to the test phase of the experiment, then this effect should also be observed when sensitization occurred after training. Rather, in Experiment 9, amphetamine pre-treated rats were as sensitive to the changed value of the reinforcer as vehicle-injected control animals. Hence the insensitivity to outcome devaluation observed in the two extinction tests in Experiment 8 can only be explained in terms of a failure to integrate knowledge about the changed value of the reinforcer with current actions rather than any differential impact of manipulations of reward value in sensitized animals.

A further notable feature of the current data was the failure to see higher baseline levels of responding in the amphetamine sensitized animals. Many studies have reported enhanced conditioned responding after sensitization with amphetamine (Harmer and Phillips, 1998; Harmer and Phillips, 1999; Taylor and Horger, 1999; Taylor and Jentsch, 2001; Phillips et al., 2002) and chronic treatment with dopamine antagonists has been shown to decrease operant responding (Varvel, Vann, Wise, Philibin and Porter, 2002). However, here baseline lever press rates did not differ between the two treatment groups. This may reflect the limited amount of training given, as higher rates did emerge after the reminder session in Experiment 8 and during the reacquisition test in Experiment 9. It may also be due to the dissociation between Pavlovian and instrumental conditioning seen in the current experiment. Indeed, in contrast to the recently reported disruption of goal-directed responding after cocaine sensitization in a Pavlovian paradigm (Schoenbaum and Setlow, 2004) we found that magazine entry behaviour was sensitive

to outcome devaluation. This finding is, however, entirely consistent with previous reports that suggest that magazine entry behaviour is under different psychological and neural control to the performance of lever pressing (Holland, 1979; Dickinson et al., 2000; Corbit et al., 2001; Corbit and Balleine, 2003) and accords with findings that habitual lever pressing but not magazine approach behaviour is insensitive to outcome devaluation (Holland, 1998; Holland, 2005; Killcross and Coutureau, 2003; Experiments 1-7). A consideration of the proximity of these response classes to reward delivery may provide a possible reason for this dissociation. In line with current findings, evidence suggests (Balleine et al., 1995) that responses proximal to the goal (such as magazine entry) remain more sensitive to motivational shifts and devaluation procedures than responses more distal to reward (such as lever pressing). This has been characterized as reflecting the development of hierarchical chains of actions whereby the performance of distal actions lead to proximal responses but only the proximal action forms a direct association with the reward (Balleine et al., 1995; Killcross & Blundell, 2002), or as reflecting a greater control of magazine approach by Pavlovian, as opposed to instrumental, contingencies (Balleine et al., 1995).

The finding that simple exposure to amphetamine renders instrumental responding insensitive to outcome devaluation concurs with evidence reviewed earlier for a neural dissociation between a goal-directed action system and a habit system. The present results also provide unequivocal evidence of dopaminergic involvement in the transformation of actions into habits. As such, these findings complement previous research implicating dopamine in the formation of S-R habits (e.g. Robbins et al., 1990;

Faure et al., 2005). As reviewed in the general introduction, the neural substrates of habit learning include the dorsolateral striatum (e.g. Yin et al., 2004), an area that receives rich innervation from midbrain dopaminergic pathways originating in the substantia nigra (Graybiel, 1990). The effects of psychostimulants on dopamine transmission within the ventral tegmental area and the nucleus accumbens are well documented (for a review see Vanderschuren and Kalivas, 2000) but evidence also suggests profound effects of psychostimulant sensitization in dorsal striatal terminal regions (Di Chiara and Imperato, 1988; Barrot et al., 1999; Canales and Graybiel, 2000; Letchworth, Nader, Smith, Friedman and Porrino, 2001; Ito, Dalley, Robbins and Everitt, 2002; Li, Kolb and Robinson, 2003; Porrino, Lyons, Smith, Daunais and Nader, 2004). The 'spiralling loop' circuitry of the striatum, whereby each striatal subregion influences its own dopamine innervation and that of its adjoining subregion in a ventral to lateral progression, may underlie these effects (Haber *et el.*, 2000). Thus through its ability to increase striatal dopamine levels, amphetamine may enhance the normal incremental process whereby S-R associations come to mediate instrumental performance. Indeed, the finding in Experiment 9 that post-training amphetamine sensitization failed to impact upon animals' susceptibility to outcome devaluation, suggests that the deficits reported in Experiment 8 are due to the accelerated 'stamping-in' of S-R associations necessary for the acquisition and maintenance of procedural habit learning. Indeed, it has been demonstrated that sensitization with psychostimulants brings about differential changes in the responsiveness of neurons in matrix and striosome compartments of the striatum; matrix activity decreases, leading to preferential activation of the striosomal system (Canales et al., 2002; Saka et al., 2004). Following recent suggestions (e.g. Canales, 2005), this shift

in activity patterns may represent the normal shift in neural activation during the transition between goal-directed and habitual behaviour, a process that is facilitated by sensitization.

CHAPTER 5

Experiments 10-13: The effect of dopamine antagonists on sensitivity to outcome devaluation after limited training in amphetamine pre-treated (Experiments 10-12) and drug-naïve animals (Experiment 13).

In the previous chapter, it was demonstrated that pre-training exposure to amphetamine led to accelerated habit formation after limited training as indexed by insensitivity to changes in reward value. Amphetamine, an indirect dopamine agonist, increases extracellular dopamine levels by blocking the actions of the dopamine transporter DAT thus preventing the removal of dopamine from the synaptic cleft. As such, the finding of enhanced habit formation after amphetamine pre-treatment suggests that dopamine transmission is critical to the development of behaviour dominated by S-R habits. The current experiments sought to provide further evidence for the involvement of dopamine in the transfer of behavioural control from goal-directed actions to habits by assessing the impact of antagonism of dopaminergic systems in an instrumental reinforcer devaluation task.

Significantly, the physiological actions of dopamine are mediated by at least five different G-protein coupled receptor subtypes, which are classified into D_1 - like and D_2 - like families. D_1 and D_2 receptor subtypes are characterised on the basis of their stimulatory or inhibitory activity on adenylyl cyclase (Kebabian and Calne, 1979) and are implicated in the great majority of dopamine-dependent behaviours. Both receptor

subtypes are widely distributed throughout the brain, with the greatest concentration within the striatum (Boyson, McGonigle and Molinoff, 1986; Wamsley, Gehlert, Filloux and Dawson, 1989), but as yet the distinct contribution to instrumental learning of these two receptors subtypes is poorly understood. However, the finding that D_1 and D_2 receptor subtypes are differentially involved in long-term potentiation (LTP) and depression (LTD) within the striatum would suggest that these receptor subtypes may have distinct learning and memory functions (Centonze, Picconi, Gubellini, Bernardi and Calabresi, 2001). For example, long-term potentiation is blocked by D_1 antagonists (Kerr and Wickens, 2001) but is facilitated by D_2 antagonists and in D_2 receptor knock-out mice (Calabresi et al., 1997; Yamamoto et al., 1999). Thus one of the aims of the current experiment was to examine the receptor subtype specificity of the sensitization effect observed in Chapter 4. The administration of various dopaminergic antagonists during training would assess whether these drugs can restore normal goal-directed behaviour in animals with pre-training exposure to amphetamine and highlight the receptor subtypes involved in the accelerated habit formation induced by amphetamine sensitization reported in the previous chapter. Evidence to date suggests that concurrent activation of both D_1 and D_2 receptor subtypes is required for the expression of repetitive and stereotypic response patterns characteristic of sensitization with psychostimulants as well as for the changes in circuit level neuroplasticity within the striatum associated with repeated drug exposure (Capper-Loup et al., 2002; Canales et al., 2002). However, their role in S-R habit learning has not been examined.

Furthermore, the administration of dopamine antagonists in non-sensitized animals would provide insights into the more general role of dopamine receptor subtypes in mediating instrumental performance. Previous work has shown that dopamine antagonists generally attenuate the rate of instrumental responding (e.g. Varvel et al., 2002) but in that study no dissociations at the receptor subtype level were made and nor was the associative structure of any behaviour probed. However, there is some evidence that D_1 and D_2 antagonists have dissociable effects on learning (e.g. Beninger and Miller, 1998). For example Eyny and Horvitz (2003) showed that the D_1 antagonist SCH23390 attenuated while the D_2 antagonist raclopride facilitated learning in a Pavlovian conditioned approach paradigm and a similar dissociation between D_1 and D_2 receptor manipulations has been reported in an odour discrimination task (Yue, Cleland, Pavlis and Linster, 2004).

In the present experiments animals were treated with the non-selective dopamine antagonist α-flupenthixol, the selective D_1 antagonist SCH 23390 or the selective D_2 antagonist eticlopride, prior to instrumental training in a reinforcer devaluation task. The drugs, doses and route of administration were selected on the basis of their effects in an instrumental conditional discrimination task developed in this laboratory (Dunn, 2003; Dunn, Futter, Bonardi and Killcross, 2005). The effect of these dopaminergic antagonists was assessed in animals sensitized with amphetamine (Experiments 10-12) or in non-sensitized drug-naïve animals (Experiment 13).

Materials and Methods

Subjects

Male Lister hooded rats were used in these experiments (Experiment 10 n = 32; Experiment 11 n = 32; Experiment 12 n = 32; Experiment 13 n = 64; Harlan UK Ltd., Bicester, Oxon, UK). At the start of behavioural test animals weighted between 263 g and 389 g. Animal care and husbandry were as described elsewhere.

Drugs

Experiments 10-12 and 13. The sensitizing d-amphetamine (Sigma-Aldrich, UK) injections (Experiments 10-12) were prepared and administered at the same dose (2mg/kg) and route (i.p.) as for the sensitization experiments described in Chapter 4. As in Chapter 4, PBS served as the vehicle control for the non-sensitized animals in Experiment 13. The amphetamine challenge injections were prepared and administered in the same manner as Chapter 4.

Experiment 10. α-Flupenthixol was dissolved in 0.9% physiological saline and administered i.p. 20 minutes prior to instrumental conditioning at a dose of 0.3mg/kg. Saline served as vehicle control and was administered at the same time, route and dose as α-flupenthixol.

Experiment 11. SCH23390 (Sigma-Aldrich, UK) was dissolved in 0.9% physiological saline and administered i.p. 15 minutes prior to instrumental conditioning at a dose of 0.005mg/kg. Saline was used as the vehicle control as above.

Experiment 12. Eticlopride (Sigma-Aldrich, UK) was dissolved in 0.9% physiological saline and administered i.p. 15 minutes prior to instrumental conditioning at a dose of 0.05mg/kg. Saline served as the vehicle control as above.

Expeiment 13. α-Flupenthixol, SCH 23390 and eticlopride were administered at the same doses, routes and times as above except that the eticlopride dose was reduced to 0.02mg/kg. Saline at a volume of 0.3ml/kg was administered to the vehicle controls.

Apparatus

The experiments were conducted in the chambers described previously in Chapter 2 and 4. The same reinforcers as in the sensitization experiments of Chapter 4 were used.

Sensitization.

In Experiments 10-12 all animals were sensitized with amphetamine. The same sensitization protocol was employed as Experiment 8 of Chapter 4. Animals in Experiment 13 underwent the same procedure except that they received 2ml/kg i.p. injections of PBS.

Behavioural training

After the sensitization treatment, animals were given one week of recovery and food withdrawal prior to the commencement of behavioural training. Following the sensitization procedure each animal was assigned to one of the eight conditioning chambers, and thereafter was always trained in that chamber. At the start of each session, the house light came on and remained on throughout the session. The house light went

out at the end of each session. Behavioural training consisted of three stages: magazine training, instrumental training and devaluation by LiCl.

Magazine training. Animals received 2 days of magazine training as described previously.

Lever press training and administration of dopamine antagonists. As in Chapter 4, animals received 2 sessions on a CRF, earning 25 rewards in each, followed by 3 sessions of RI30 lever press training, earning 40 rewards in each session and hence a total of 120 rewards on the RI30 schedule. However, prior to each of these lever press training sessions, animals received an i.p. injection of a dopamine antagonist (Drug groups) or the equivalent volume of control vehicle solution (Control group). In Experiments 10-12 half the animals (group Drug) received injections of a dopamine antagonist (α-flupenthixol in 10, SCH23390 in 11 and eticlopride in 12) and the other half (Controls) injections of saline. In Experiment 13, 16 animals were administered with α-flupenthixol, 16 with SCH23390, 16 with eticlopride and 16 served as vehicle-injected controls.

Devaluation by lithium chloride

Taste aversion training. After the final day of instrumental lever press training, animals received three days of reward devaluation training with LiCl. This was conducted in the same manner as described in Chapter 4 with half the animals received pairings of the instrumental outcome with LiCl (devalued group) and the other half with saline (non-devalued group). Taste aversion training was conducted drug-free.

Extinction test. 24 hours after the final session of taste aversion training, animals received a 10-minute, drug-free extinction test conducted in the absence of reward delivery. This was carried out as previously described.

Reacquisition test. In order to confirm that the taste aversion procedure had successfully devalued the outcome for the devalued group, all animals underwent a 15 minute, drug-free reacquisition test. This test was performed as described in Chapter 4.

Consumption test (Experiment 12 only). Animals were placed in feeding cages and given unrestricted access to the instrumental outcome for fifteen minutes. The test was conducted drug-free.

Activity assay. To confirm sensitization, all animals were subject to a 0.5mg/kg (*i.p.*) amphetamine challenge injection and locomotor activity was assessed for 30 minutes in the same activity boxes and in the same way as Chapter 4.

Data analysis. Statistical analysis was performed using analysis of variance (ANOVA) with between subject factors of devaluation (devalued versus non-devalued) and drug treatment (either dopamine antagonist or saline). As the standard deviation was proportional to the mean, the extinction data were subject to logarithmic transformations (Howell, 2002). This was conducted as described for Experiment 5 in Chapter 3. Significant main effects with more than 2 levels were explored with Tukey post-hoc tests.

Results

Experiment 10. The effect of α-flupenthixol on sensitivity to outcome devaluation after limited training in animals pre-treated with amphetamine.

Instrumental training. By the end of the three days of RI30 training, all animals had acquired the instrumental response and achieved a stable level of responding. Figure 5.1.1 displays lever press per minute for the final session and it is clear from this figure that α-flupenthixol treatment produced overall lower rates of responding compared to sensitized animals treated with saline. This was confirmed statistically by a main effect of drug ($F_{(1,28)} = 7.98$, $p<0.05$). However, as the length of each session was determined by the number of reinforcers earned (40 in each) and not time, α-flupenthixol treated animals obtained the same number of reinforcers (120) as controls and hence any differential sensitivity to outcome devaluation observed in the subsequent extinction test cannot be accounted for in terms of differential exposure to the reinforcer. As the critical comparisons at test are between devalued and non-devalued groups within each drug group, it is unlikely that any differences in sensitivity to outcome devaluation are due to these baseline differences. Significantly in this respect, there was no effect of intended devaluation ($F<1$) nor interaction between drug and devaluation ($F<1$). In contrast to the depressive effects of α-flupenthixol on lever press acquisition, there was no effect of drug on magazine entry behaviour. ANOVA yielded no effect of drug ($F<1$) or devaluation ($F_{(1,28)} = 2.224$, p= .145), and no interaction ($F<1$) (mean magazine entries per minute (±SD) saline to-be-devalued group = 4.4 (±1.5); saline to-be-valued group = 6.6 (±5.2);

α-flupenthixol to-be-devalued group = 4.0 (±2.9); α-flupenthixol to-be-valued group = 5.3 (±2.6)).

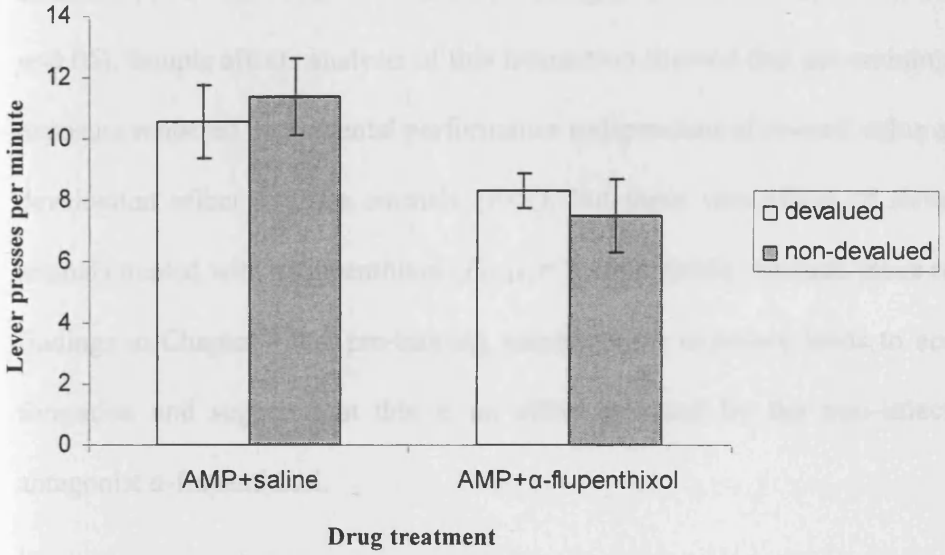

Figure 5.1.1 Mean lever presses per minute (±SEM) on the final day of acquisition in amphetamine-sensitized animals administered saline or α-flupenthixol during training.

Extinction test – lever press performance. In order to take account of baseline differences and reduce within subject variability in ANOVA, lever press performance in the extinction test is presented as a proportion of baseline responding. These are presented in Figure 5.1.2. The suggestion from this figure is that administration of α-flupenthixol during training (group AMP+α-flupenthixol) restored goal-sensitivity as the animals in the devalued group (white bars) showed a selective depression in lever press rates compared to animals in the non-devalued group (grey bars). On the other hand the responding of animals exposed to amphetamine before training but administered saline

during training (group AMP+saline) appeared to be impervious to the current value of the reinforcer as shown by equivalent rates across the two devaluation groups. There was no effect of drug ($F<1$) but there was a main effect of devaluation ($F_{(1,28)} = 6.6$, $p<0.05$) and critically ANOVA revealed a significant drug x devaluation interaction ($F_{(1,28)} = 4.3$, $p<0.05$). Simple effects analysis of this interaction showed that pre-training amphetamine exposure rendered instrumental performance independent of reward value as there was no devaluation effect in these animals ($F<1$), but there was effect of devaluation in the animals treated with α-flupenthixol ($F_{(1,14)} = 7.15$, $p<0.05$). As such these results replicate findings in Chapter 4 that pre-training amphetamine exposure leads to accelerated habit formation and suggest that this is an effect reversed by the non-selective dopamine antagonist α-flupenthixol.

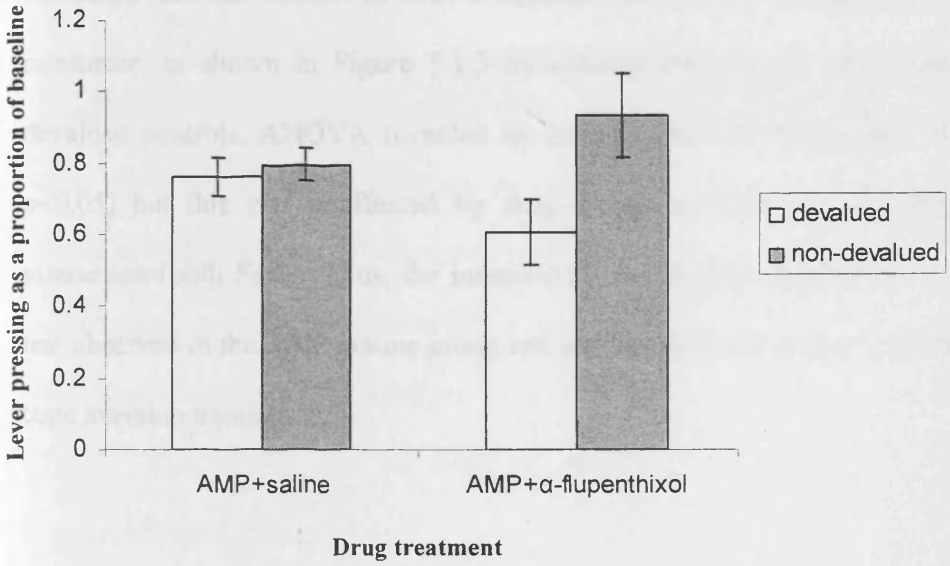

Figure 5.1.2 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after devaluation by LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Extinction test – magazine entry behaviour. Analysis of magazine entry behaviour during the extinction test revealed a main effect of devaluation ($F_{(1,28)} = 12.84$, $p<0.05$) but no effect of drug or interaction between these factors (both Fs<1). (Mean magazine entries per minute as a proportion of baseline (±SD) saline devalued group = 0.3 (±0.19); saline non-devalued group = 1.2 (±0.78); α-flupenthixol devalued group = 0.48 (±0.38); α-flupenthixol non-devalued group = 1.0 (±0.66)). Thus in contrast to lever press performance, magazine entry behaviour was sensitive to outcome value irrespective of drug and suggests that the LiCl treatment successfully devalued the value of the instrumental outcome.

Reacquisition test - Lever press performance. The results of the rewarded reacquisition test confirmed that animals in both devaluation groups had developed an aversion to the reinforcer, as shown in Figure 5.1.3 by reduced lever press rates compared to non-devalued controls. ANOVA revealed an overall effect of devaluation ($F_{(1,28)} = 10.04$, $p<0.05$) but this was unaffected by drug group as there was no effect of drug or interaction (both Fs<1). Thus, the insensitivity to outcome devaluation in the extinction test observed in the AMP+saline group can not be attributed to any differential impact of taste aversion training.

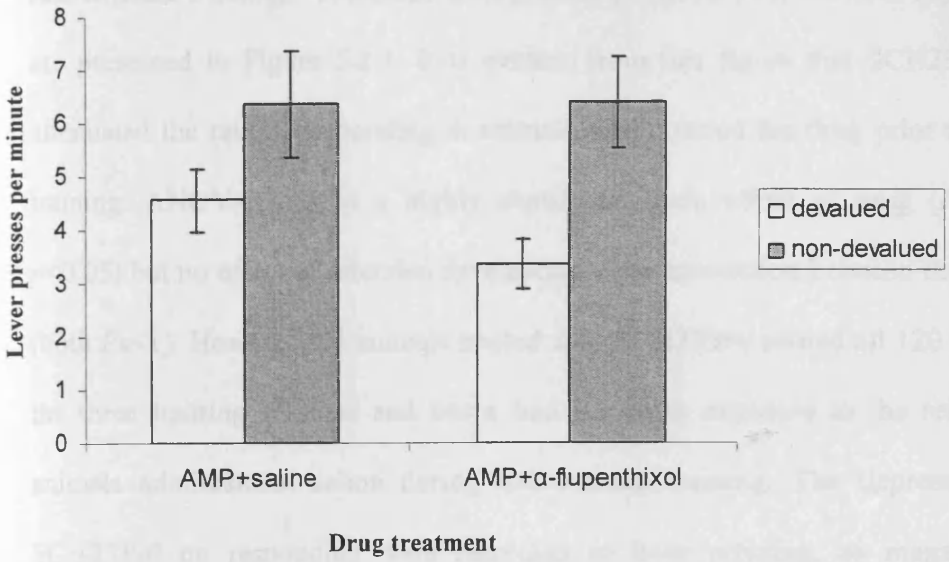

Figure 5.1.3 Mean lever presses per minute (±SEM) in the rewarded reacquisition test after devaluation with LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Reacquisition test – Magazine entry behaviour. Similarly, magazine entry behaviour during the 15 minute reacquisition test was sensitive to the changed value of the reinforcer. Both devalued groups performed considerably fewer magazine entries during the test compared to the non-devalued controls ($F_{(1,28)}$ = 11.57 p<0.05) (Mean magazine entries per minute (±SD) Devalued group = 1.912 (±1.0671); Non-devalued group = 3.098 (±2.078). There was no effect of drug ($F_{(1,28)}$ = 1.69, p=0.2) nor a drug x devaluation interaction (F<1).

Experiment 11. The effect of SCH23390 on sensitivity to outcome devaluation after limited training in animals pre-treated with amphetamine.

Instrumental training. The mean lever presses per minute for the final day of acquisition are presented in Figure 5.2.1. It is evident from this figure that SCH23390 markedly attenuated the rate of responding in animals administered the drug prior to instrumental training. ANOVA yielded a highly significant main effect of drug ($F_{(1,28)}$ = 36.39, p<0.05) but no effect of intended devaluation or an interaction between these two factors (both Fs<1). However, all animals treated with SCH23390 earned all 120 rewards across the three training sessions and hence had the same exposure to the reinforcer as the animals administered saline during instrumental training. The depressive effects of SCH23390 on responding were restricted to lever pressing, as magazine approach behaviour was unaffected by the drug (mean magazine entries per minute (±SD) saline to-be-devalued group = 5.8 (±2.4); saline to-be-valued group = 4.5 (±2.2); SCH23390 to-be-devalued group = 4.2 (±2.6); SCH23390 to-be-valued group = 4.2 (±1.8)). Statistically, there was no effect of drug ($F_{(1,28)}$ = 1.43, p=0.241), intended devaluation nor an interaction (both Fs<1).

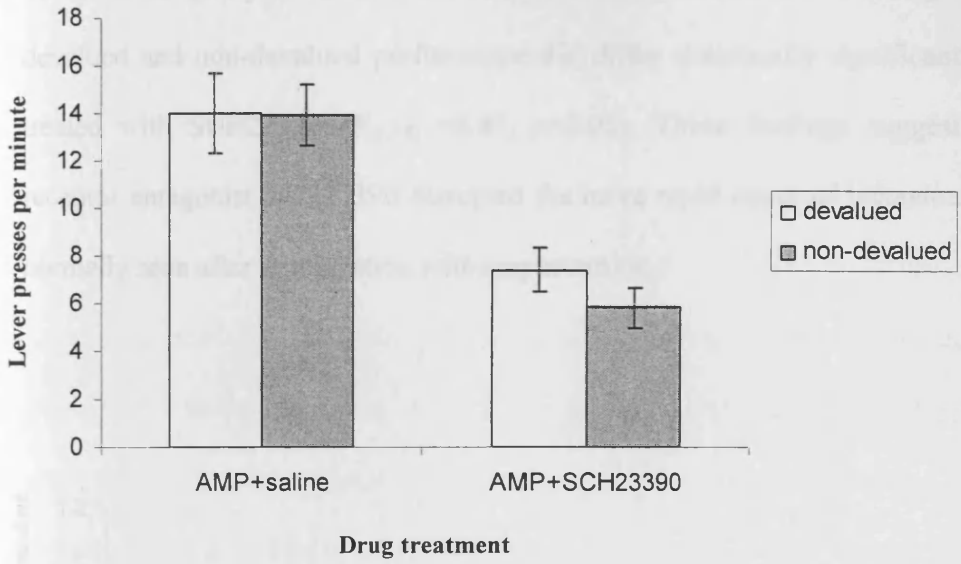

Figure 5.2.1 Mean lever presses per minute (±SEM) on the final day of acquisition of amphetamine pre-treated animals administered saline or SCH23390 during training.

Extinction test – lever press performance. The lever press performance of saline injected and SCH23390-treated group during the 10 minute extinction as a proportion of their baseline responding is presented in Figure 5.2.2. Inspection of this figure suggests that the instrumental performance of animals treated with SCH23390 during training was guided by outcome expectancy as the devalued group (white bars) performed fewer lever presses as a proportion of baseline compared to the non-devalued group (grey bars). Conversely, the responding of the amphetamine + saline group in this test was not goal-directed as demonstrated by their failure to show sensitivity to the change in reward value. This description of the data was confirmed statistically by ANOVA which revealed a main effect of devaluation ($F_{(1,28)} = 9.16$, $p<0.05$), no effect of drug ($F<1$) and significantly, a devaluation x drug interaction ($F_{(1,28)} = 7.15$, $p<0.05$). Subsequent analysis

of this interaction yielded no effect of devaluation in the AMP + saline group ($F<1$) but devalued and non-devalued performance did differ statistically significantly in animals treated with SCH23390 ($F_{(1,14)}$ =8.82, $p<0.05$). These findings suggest that the D_1 receptor antagonist SCH23390 disrupted the more rapid onset of behavioural autonomy normally seen after sensitization with amphetamine.

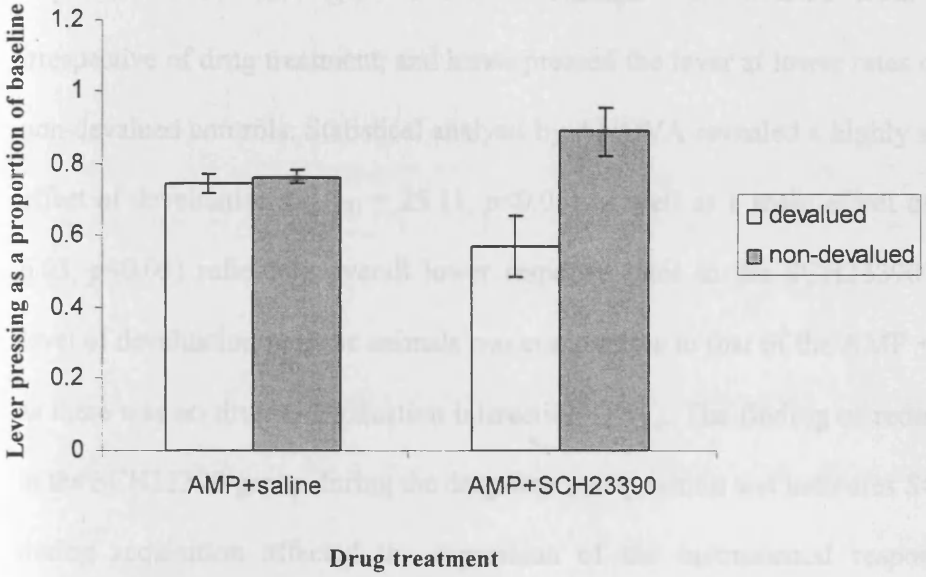

Figure 5.2.2 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after devaluation by LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Extinction test – magazine entry behaviour. In contrast, magazine performance during the extinction test was sensitive to the changed value of the reinforcer in both drug groups (mean magazine entries per minute as a proportion of baseline (±SD) saline devalued group = 0.66 (±0.28); saline non-devalued group = 1.3 (±0.87); SCH23390

devalued group = 0.33 (±0.31); SCH23390 non-devalued group = 1.5 (±0.72)). Indeed, ANOVA revealed only a main effect of devaluation ($F_{(1,28)} = 18.52$, $p<0.05$), no effect of drug ($F<1$) nor an interaction ($F_{(1,28)} = 1.59$, $p= 0.22$).

Reacquisition test - Lever press performance. The effectiveness of the taste aversion training in devaluing the instrumental outcome is further supported by analysis of lever press rates performed in the rewarded reacquisition test shown in Figure 5.2.3. The impression from this figure is that all animals were averted from the reinforcer, irrespective of drug treatment, and hence pressed the lever at lower rates compared to the non-devalued controls. Statistical analysis by ANOVA revealed a highly significant main effect of devaluation ($F_{(1,28)} = 25.11$, $p<0.05$) as well as a main effect of drug ($F_{(1,28)} = 6.03$, $p<0.05$) reflecting overall lower response rates in the SCH23390 group, but the level of devaluation in these animals was comparable to that of the AMP + saline animals as there was no drug x devaluation interaction ($F<1$). The finding of reduced responding in the SCH23390 group during the drug-free reacquisition test indicates SCH23390 given during acquisition affected the expression of the instrumental response as well as performance in training.

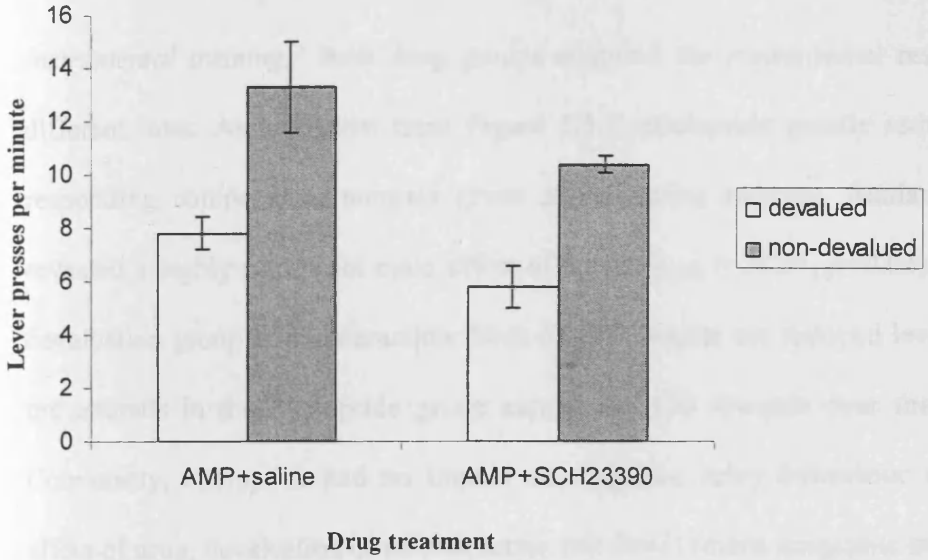

Figure 5.2.3 Mean lever presses per minute (±SEM) in the rewarded reacquisition test after devaluation with LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Reacquisition test – Magazine entry behaviour. Magazine entry behaviour was equally sensitive to outcome value in both drug groups during the reacquisition test (Mean magazine entries per minute (±SD): Devalued group = 3.983 (±3.276); Non-devalued group = 8.036 (±3.814)). Statistically, there was an overall effect of devaluation ($F_{(1,28)}$ = 10.52, p<0.05) but no effect of drug (F<1) nor an interaction ($F_{(1,28)}$ = 1.32, p=0.26). Thus in contrast to lever press performance in the reacquisition test, magazine approach behaviour was unaffected by SCH23390 and this provides yet further evidence that magazine entry behaviour is under different neural control to lever pressing.

Experiment 12. The effect of eticlopride on sensitivity to outcome devaluation after limited training in animals pre-treated with amphetamine.

Instrumental training. Both drug groups acquired the instrumental response, albeit at different rates. As is evident from Figure 5.3.1, eticlopride greatly reduced the rate of responding compared to animals given saline during training. Statistically, ANOVA revealed a highly significant main effect of drug ($F_{(1,28)} = 34.21$, $p<0.05$) but no effect of devaluation group or an interaction (both Fs<1). Despite the reduced lever press rate, all the animals in the Eticlopride group earned the 120 rewards over the three sessions. Conversely, eticlopride had no impact on magazine entry behaviour as there was no effect of drug, devaluation or an interaction (all Fs<1) (mean magazine entries per minute (±SD): Saline group = 4.311 (±1.726); Eticlopride group = 5.055 (±2.359)).

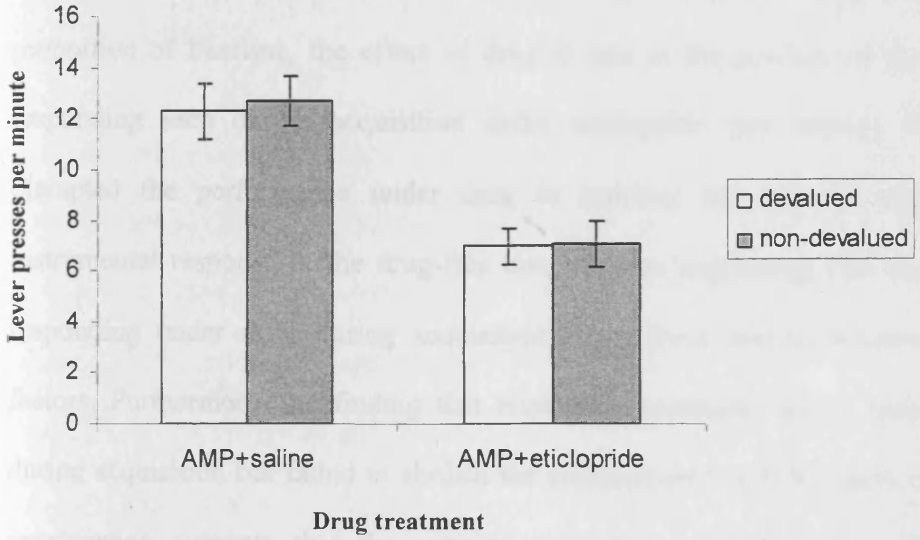

Figure 5.3.1 Mean lever presses per minute (±SEM) on the final day of acquisition of amphetamine pre-treated animals administered saline or eticlopride during training.

Extinction test – lever press performance. The mean lever presses per minute in the critical extinction test are presented in Figure 5.3.2. It is clear from this figure that none of the animals, irrespective of drug group, was sensitive to the changed value of the reinforcer as both devalued groups responded at equivalent rates to the non-devalued controls. This was confirmed statistically as there was no effect of devaluation ($F<1$) and no interaction between drug and devaluation factors ($F<1$). Eticlopride therefore failed to reverse the effect of pre-training amphetamine exposure on goal-sensitivity after limited training and responding in both groups was habitual even after limited training. However, ANOVA did reveal a highly significant main effect of drug ($F_{(1,28)} = 15.58$, $p<0.05$), reflecting overall higher rates of responding as a proportion of baseline in the eticlopride group. As the extinction test was conducted drug-free and the data were analysed as a proportion of baseline, the effect of drug at test is the product of the lower rates of responding seen during acquisition under eticlopride (see above). Thus eticlopride disrupted the performance under drug in training but not the expression of the instrumental response in the drug-free test, perhaps suggesting that the lower rates of responding under drug during acquisition arose from non-associative (e.g. motoric) factors. Furthermore, the finding that eticlopride-treatment led to reduced responding during acquisition but failed to abolish the enhancement of S-R habits by amphetamine sensitization suggests that the restoration of goal-sensitivity by α-flupenthixol and SCH23390 (Experiments 10 and 11, see above) cannot be attributed to their depressive effects on response rates during acquisition.

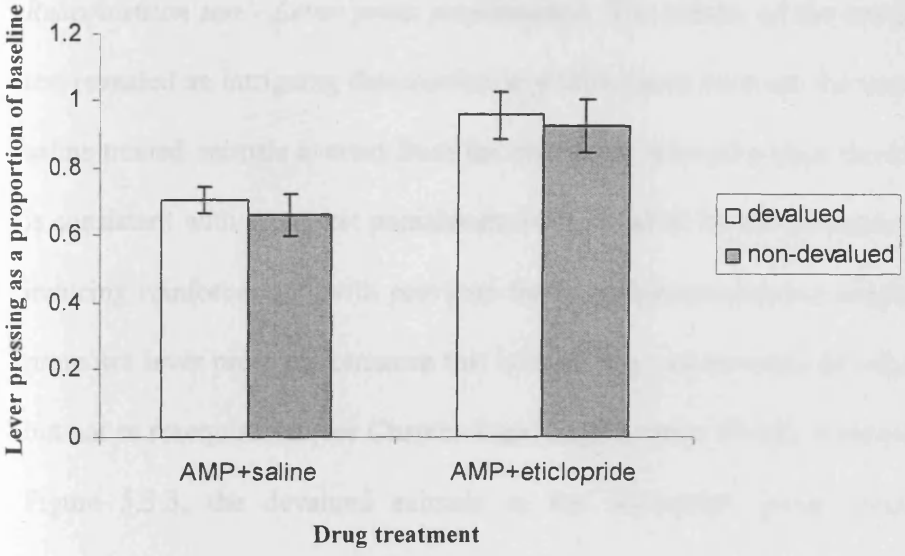

Figure 5.3.2 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after devaluation by LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Extinction test – magazine entry behaviour. Despite the insensitivity of lever pressing to outcome devaluation, magazine entry behaviour in both devalued groups was reduced compared to non-devalued controls (mean magazine entries per minute as a proportion of baseline (±SD) saline devalued group = 0.56 (±0.17); saline non-devalued group = 1.3 (±0.88); Eticlopride devalued group = 0.68 (±0.4); Eticlopride non-devalued group = 1.3 (±0.73)). Statistical analysis revealed only an effect of devaluation ($F_{(1,28)}$ = 10.58, p<0.05) and no effect of drug nor an interaction (both F's<1). Thus the demonstration that lever press performance in the extinction test was under the control of S-R habits, whereas magazine approach behaviour was guided by outcome value, indicates that the LiCl treatments must have successfully devalued the instrumental outcome.

Reacquisition test - Lever press performance. The results of the rewarded reacquisition test revealed an intriguing dissociation in performance between the two drug groups. The saline treated animals averted from the reinforcer showed a clear devaluation effect: this is consistent with the direct punishment of S-R habits by the presentation of the nausea-inducing reinforcer and with previous findings that pre-training amphetamine exposure promotes lever press performance that is insensitive to outcome devaluation in extinction but not in reacquisition (see Chapter 4 and Experiments 10-12). However as is clear from Figure 5.3.3, the devalued animals in the eticlopride group pressed the lever at comparable rates to the non-devalued controls even though responding was reinforced with the reward that had been previously paired in these animals with gastric malaise. This description of the data was supported statistically by ANOVA which revealed a main effect of devaluation ($F_{(1,28)} = 10.38$, $p<0.05$), no effect of drug ($F<1$), but crucially a significant interaction between these two factors ($F_{(1,28)} = 5.47$, $p<0.05$). Subsequent analysis of this interaction with simple effects confirmed that saline-treated animals had acquired an aversion to the reinforcer and could use this representation to guide instrumental performance when presented with the consequences of their actions in reacquisition as there was a highly significant effect of devaluation in these animals ($F_{(1,14)} = 12.17$, $p<0.05$). There was no such effect in the eticlopride-treated animals ($F<1$). This can be taken as evidence that instrumental performance in eticlopride treated animals was completely impervious to reward value and had become compulsive. However, it is possible that this insensitivity arose from a failure of the taste aversion training.

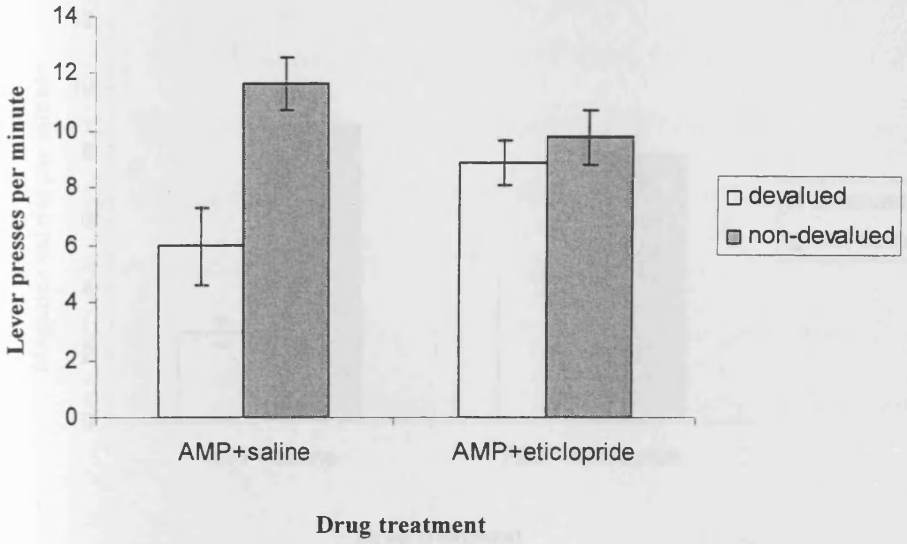

Figure 5.3.3 Mean lever presses per minute (±SEM) in the rewarded reacquisition test after devaluation with LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Reacquisition test – Magazine entry behaviour. Significantly, analysis of magazine entry behaviour during the rewarded reacquisition test suggests that all animals, regardless of drug treatment, had acquired an aversion to the reinforcer. The mean magazine entries per minute in this test are displayed in Figure 5.3.4 and in stark contrast to the lever press data reviewed above, magazine approach behaviour was sensitive to reward value in both drug groups. ANOVA yielded no effect of drug ($F<1$) and a highly significant effect of devaluation ($F_{(1,28)} = 45.6$, $p<0.05$). The suggestion from Figure 5.3.4 is that the devaluation effect may have been slightly attenuated in the eticlopride group but there was no statistical evidence for this as the interaction failed to reach significance ($F_{(1,28)} = 2.74$, $p=0.11$).

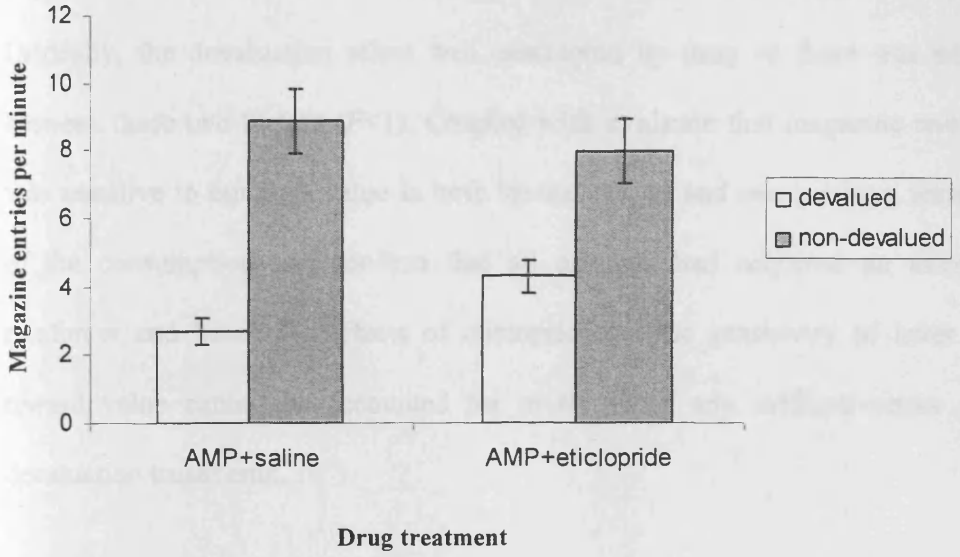

Figure 5.3.4 Mean magazine entries per minute (±SEM) in the rewarded reacquisition test after devaluation with LiCl (devalued - white bars) or no devaluation (non-devalued grey bars).

Consumption test. In order to confirm that the differential sensitivity of lever press to reward value observed in the reacquisition test could not be explained in terms of any failure of eticlopride animals to acquire an aversion to the reinforcer, all animals were given free access to the instrumental outcome and consumption was measured over a 15-minute period. Results of this consumption test revealed that all animals averted from the reinforcer consumed statistically significantly less of the instrumental outcome compared to the non-devalued controls (mean consumption in ml (±SD): Devalued saline group = 3.3 (±1.6); Devalued eticlopride group = 4.2 (±2.0); Non-devalued saline group = 6.2 (±2.7); Non-devalued eticlopride group = 7.00 (±1.7)). ANOVA revealed a main effect of devaluation ($F_{(1,28)}$ =15. 78, $p<0.05$) and a non-significant trend towards marginally

higher overall consumption in eticlopride-treated animals ($F_{(1,28)}$ = 1.39, p=0.25). Critically, the devaluation effect was unaffected by drug as there was no interaction between these two factors (F<1). Coupled with evidence that magazine entry behaviour was sensitive to outcome value in both the extinction and reacquisition tests, the results of the consumption test confirm that all animals had acquired an aversion to the reinforcer and hence the effects of eticlopride on the sensitivity of lever pressing to reward value cannot be accounted for in terms of any ineffectiveness of the LiCl devaluation treatments.

Experiment 13. The effect of α-flupenthixol, SCH23390 and eticlopride on the sensitivity to outcome devaluation after limited training in drug-naïve animals.

Instrumental training. The lever press rates on the final day of acquisition are presented in Figure 5.4.1. It is clear from this figure that in line with predictions, the dopamine antagonists reduced the rate of responding and this effect was particularly marked in animals treated with SCH23390 and eticlopride. Despite the reduction in the rate of responding, all animals earned 120 reinforcers across the three sessions. This description of the data was confirmed by ANOVA which revealed a main effect of drug ($F_{(3,56)}$ =10.65, p<0.05) but no effect of intended devaluation nor an interaction between these factors (both F's<1). Subsequent post-hoc analysis with Tukey tests confirmed that both SCH23390- (p<0.05) and eticlopride- (p<0.05) treated animals responded at lower rates than saline treated animals. However magazine entry behaviour was unaffected by any of these factors as there was no effect of drug, intended devaluation or interaction (highest

$F_{(1,56)}$ =1.72, p=0.2) (Mean magazine entries per minute (±SD): Saline group = 5.020(1.947); α-flupenthixol group = 4.605 (±2.084); SCH23390 group = 4.290 (±2.684); Eticlopride group = 3.567 (±1.887)).

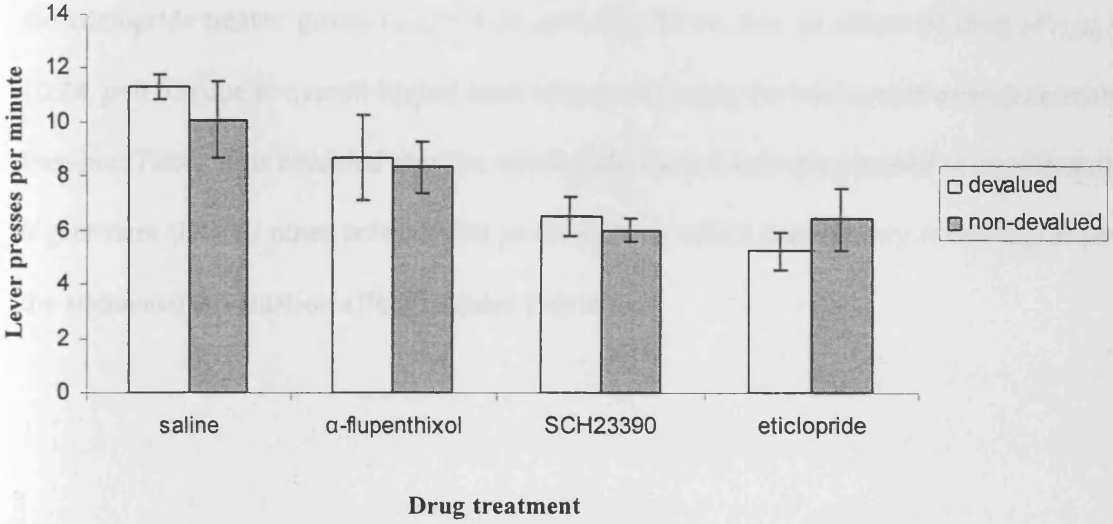

Figure 5.4.1 Mean lever presses per minute (±SEM) on the final day of acquisition in animals administered saline or α-flupenthixol or SCH23390 or eticlopride during training.

Extinction test – lever press performance. Figure 5.4.2 displays the lever press performance in the extinction test following devaluation by LiCl. Inspection of this figure suggests that saline controls and animals given SCH23390 and α-flupenthixol during training were goal-directed as animals averted from the reinforcer showed a marked suppression in lever press performance compared to non-devalued control animals. The suggestion from this figure is that the devaluation effect may have been attenuated in animals treated with eticlopride., An ANOVA revealed only a main effect

of devaluation ($F_{(1,56)}$ =22.95, p<0.05) and no interaction between drug and devaluation (F<1). Nevertheless planned pairwise comparisons confirmed a robust devaluation effect in the saline-treated ($_{t(14)}$ = 3.08, p<0.05), SCH23390-treated ($_{t(14)}$ = 4.03, p<0.05) and α-flupenthixol-treated ($_{t(14)}$ = 4.03 p<0.05) animals but there was no effect of devaluation in the eticlopride treated group ($_{t(14)}$ = 1.21 p=0.25), There was an effect of drug ($F_{(3,56)}$ = 10.28, p<0.05) due to overall higher rates of responding in the eticlopride treated animals. Post-hoc Tukey tests revealed that the eticlopride-treated animals pressed at significantly higher rates than all other animals (all ps<0.001); an effect presumably reflecting in part the attenuated devaluation effect in these animals.

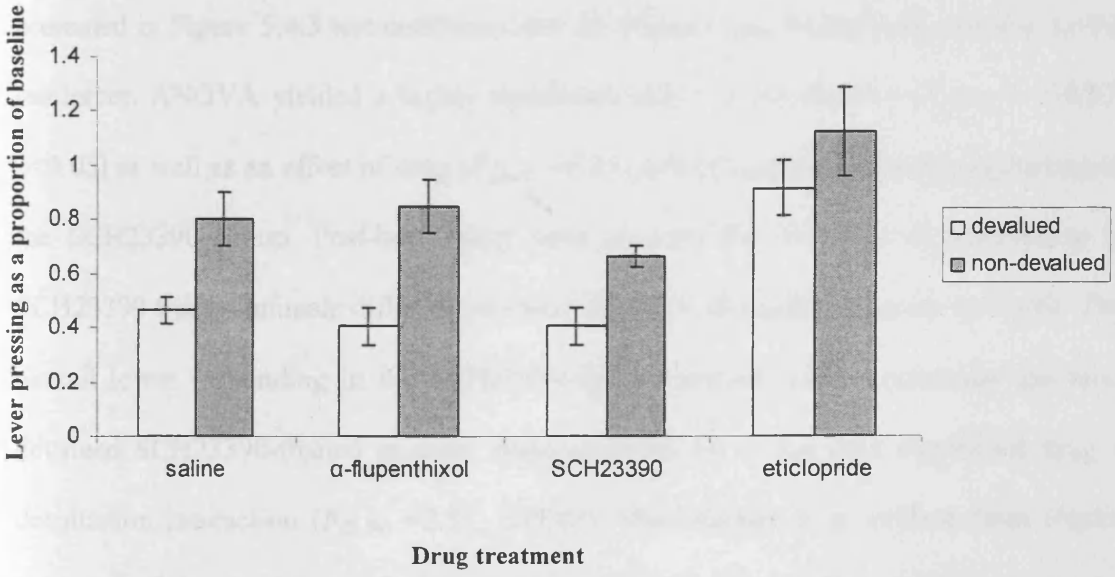

Figure 5.4.2 Mean lever presses per minute as a proportion of baseline (±SEM) in the extinction test after devaluation by LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Extinction test – magazine entry behaviour. Analysis of magazine entry behaviour during the 10-minute extinction test suggests that the LiCl treatment successfully devalued the outcome for all animals as there was a main effect of devaluation ($F_{(1,56)}$ = 7.0, $p<0.05$) (mean magazine entries per minute as a proportion of baseline (±SD) devalued group = 0.44 (±0.33); non-devalued group = 0.77 (±0.47)). There was a marginal effect of drug ($F_{(3,56)}$ =2.26, p=0.052) as there were overall higher magazine entries in the SCH22390 group. Nevertheless, drug did not impact on the level of devaluation seen in any of the animals as there was no drug x devaluation interaction ($F_{(3,56)}$ =1.53, p=0.22).

Reacquisition test - Lever press performance. The results of the rewarded reacquisition presented in Figure 5.4.3 test confirmed that all animals had acquired an aversion to the reinforcer. ANOVA yielded a highly significant effect of devaluation ($F_{(1,56)}$ = 138.83, $p<0.05$) as well as an effect of drug ($F_{(3,56)}$ =2.77, $p<0.05$) reflecting lower responding in the SCH23390 group. Post-hoc Tukey tests showed that the rate of responding in SCH23390 treated animals differed only from that of α-flupenthixol group ($p<0.05$). The overall lower responding in the SCH23390-treated animals and in particular the non-devalued SCH23390-treated animals, would account for a marginal significant drug x devaluation interaction ($F_{(3,56)}$ =2.51, p=0.07). Nevertheless it is evident from Figure 5.4.3 that all devalued groups had acquired a robust aversion to the instrumental outcome and consequently suppressed lever press responding during the rewarded test.

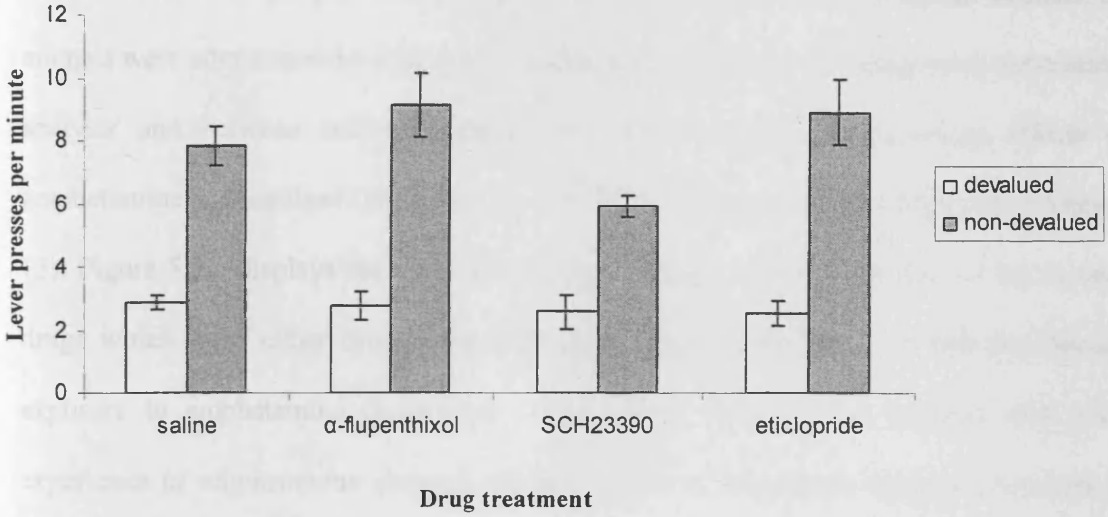

Figure 5.4.3 Mean lever presses per minute (±SEM) in the rewarded reacquisition test after devaluation with LiCl (devalued - white bars) or no devaluation (non-devalued - grey bars).

Reacquisition test – Magazine entry behaviour. This impression was also confirmed by analysis of magazine approach behaviour during the rewarded reacquisition test, with all animals in the devalued groups performing fewer magazine entries compared to the non-devalued controls ($F_{(1,56)}$ = 28.01, p<0.05) (mean magazine entries per minute (±SD) devalued group = 1.7 (±1.5); non-devalued group = 3.8 (±1.9)). There was also a main effect of drug ($F_{(3,56)}$ =6.52, **p**<0.05) as the α-flupenthixol treated animals had higher rates of magazine approach behaviour (p<0.05) but this heightened responding did not impact on sensitivity of magazine entry behaviour to outcome devaluation as there was no drug x devaluation interaction (F<1).

Activity Assay

In order to confirm the presence of sensitization in amphetamine pre-treated animals, all animals were administered a 0.5mg/kg amphetamine challenge allowing cross experiment analysis and between subject comparisons of the locomotor activating effects of amphetamine in sensitized (Experiments 10-12) and non-sensitized animals (Experiment 13). Figure 5.5.1 displays the mean photobeam breaks in the activity test of the various drugs which were either drug-naive (non-sensitized – white bars) or had pre-training exposure to amphetamine (sensitized – grey bars). As expected animals with prior experience of amphetamine showed elevated levels of locomotor activity compared to drug-naïve animals. ANOVA with between-subject factors of sensitization (sensitized with amphetamine or non-sensitized drug-naïve animals) and drug administered during training (saline, α-flupenthixol, SCH23390 or eticlopride) yielded a highly significant effect of sensitization ($F_{(1,144)} = 48.91$, $p<0.05$) but also an effect of drug ($F_{(3,144)}= 4.8$, $p<0.05$) due to higher locomotor activity in response to the amphetamine challenge in all animals treated with eticlopride during. There was, however, no interaction between sensitization and drug ($F_{(3,144)} = 1.7$, $p=0.17$). These results confirm that the amphetamine pre-treatment had successfully sensitized animals to amphetamine and provide indirect evidence that antagonism with the D_2 antagonist eticlopride enhances the locomotor activating effects of amphetamine irrespective of prior experience with the amphetamine.

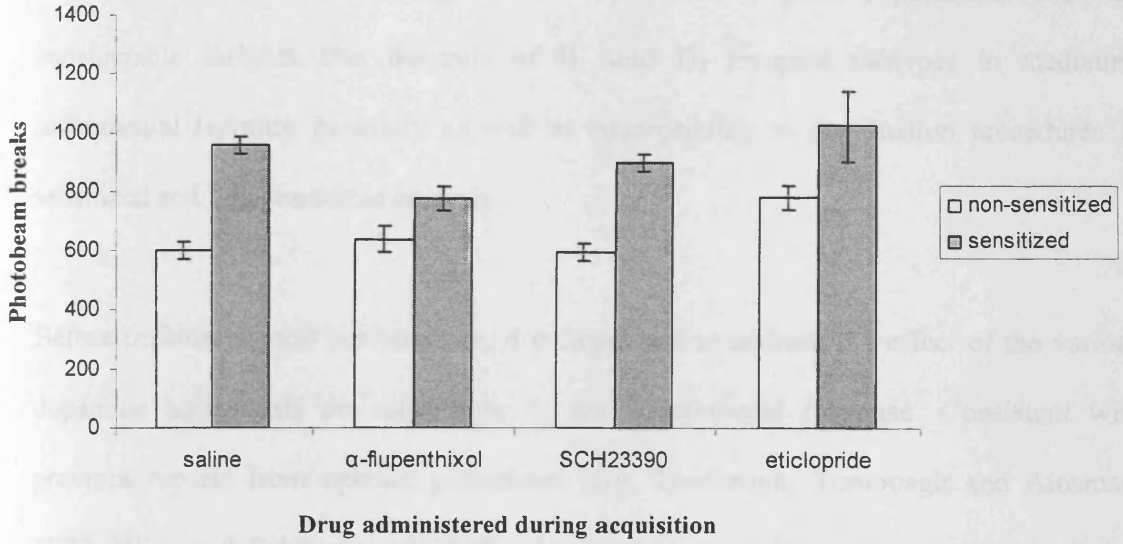

Figure 5.5.1 Mean total photobeam breaks in the activity test following an amphetamine challenge injection administered to drug-naïve animals (non-sensitized - white bars) and animals with pre-training amphetamine exposure (sensitized – grey bars).

Discussion

The experiments reported here examined the effects of both non-selective and selective dopamine antagonists on instrumental performance in a reinforcer devaluation task either in animals given pre-training exposure to amphetamine or in drug naïve animals. Significantly, the experiments replicated the findings of Chapter 4 that pre-training exposure to amphetamine renders instrumental performance autonomous of the current value of the reinforcer even after limited training. The results demonstrated that accelerated habit formation seen after amphetamine sensitization is reversed by D_1, but

enhanced by D_2 receptor antagonists. Furthermore, these experiments provided considerable insights into the role of D_1 and D_2 receptor subtypes in mediating instrumental learning generally as well as susceptibility to devaluation procedures in sensitized and non-sensitized animals.

Before considering test performance, it is important to address the effect of the various dopamine antagonists on acquisition of the instrumental response. Consistent with previous reports from operant procedures (e.g. Tombaugh, Tombuagh and Ainsman, 1979; Wise and Schwarts, 1981) the administration of dopamine antagonists during training severely retarded the rate of acquisition in both sensitized and non-sensitized animals. However, in the current experiments a dissociation at the receptor subtype level was found between the performance of instrumental responses under drug and their expression in drug-free tests. As dopamine has been implicated in various non-associative factors such as motivation, attention and sensorimotor control that contribute to learning, any effects of dopamine antagonism that are restricted to the performance of an instrumental response under drug can be attributed to these non-associative factors. However, if effects of dopaminergic manipulations are seen on the drug-free expression of learned instrumental responses, for example in the current experiments in the extinction and reacquisition tests, then this can be taken as evidence that a dopaminergic agent has actually modulated the course of associative learning.

Here, the non-selective antagonist α-flupenthixol and the selective D_2 antagonist eticlopride reduced the rate of instrumental responding during acquisition, but no effects

were observed at test. The finding that these animals showed comparable rates of responding to saline controls in the drug-free reacquisition test indicate that these drugs disrupted the performance but not the acquisition and subsequent expression of instrumental conditioning and as such suggest that their effects on operant responding are via non-associative factors. Furthermore, animals administered eticlopride during training displayed elevated levels of responding as a proportion of baseline in the extinction test (both in animals sensitized with amphetamine in Experiment 12 and amphetamine naïve animals in Experiment 13). This recovery of responding in the drug-free extinction text indicates that these animals learnt the instrumental response at the same rate as saline controls and that antagonism of D_2 receptors therefore only disrupted the performance of that response during acquistion. Dopamine activity in the nucleus accumbens, a structure containing the highest concentration of D_2 receptors in the rat brain (Bentivoglio and Morelli, 2005), has been widely implicated in the reinforcing and motivational properties of both natural rewards and dugs of abuse (e.g. Hernandez and Hoebel, 1988; Mark, Smith, Rada and Hoebel, 1994; Wyvell and Berridge, 2000). Thus the disruptive effects of agents selectively and non-selectively targeting D_2 receptors on the performance but not the acquisition of the instrumental response may have been due to decreased motivation associated with these drugs. However, the reduced rate of responding could have equally arisen as a result of the profound motor impairments typically produced by D_2 antagonists (e.g. Fowler and Liou, 1994). Whether the disruption was caused by motivational or motor factors or a combination of the two, it is clear that D_2 receptor antagonism impaired the performance but not the learning of the instrumental response in the current experiments. Conversely, D_1 antagonism by

SCH23390 not only affected the performance during training but it also reduced the expression of learned instrumental responses at test. Indeed, animals administered the D_1 antagonist during training displayed lower rates of responding in the drug-free reacquisition test compared to saline controls and other drug groups. The test was conducted five days after the last SCH23390 treatment and hence the reduced rate of responding cannot be accounted for solely in terms of drug induced motivational or sensorimotor deficits. The results are consistent with previous reports of disruption to operant responding by SCH23390 (e.g. Nakajima, 1986; Sharf, Lee and Ranaldi, 2005) and suggest that D_1 receptors are involved in the associative learning underpinning instrumental responding.

In stark contrast to the effects of dopaminergic drugs on instrumental performance, antagonism of dopaminergic systems failed to impact on magazine approach behaviour (but see Choi, Balsam and Horvitz,, 2005). Both during acquisition and at test there was no effect of the various dopaminergic agents used in the current experiments on magazine entry behaviour. Furthermore in line with evidence from Chapter 4, magazine approach behaviour remained sensitive to outcome devaluation even when instrumental performance (see below) was impervious to changes in reward value. Thus the deficits in instrumental performance observed cannot *simply* be attributable to motoric dysfunction as any drug induced motor impairment would presumably impact on magazine approach behaviour as well as lever pressing. To the extent that magazine approach behaviour in a free operant procedure depends on Pavlovian contingencies, these findings provide yet further evidence that Pavlovian and instrumental conditioning are subserved by distinct

psychological and neural processes (e.g. Holland, 1979; Dickinson et al., 2000; Corbit et al., 2001).

As expected, animals that were not exposed to amphetamine prior to training showed normal sensitivity to outcome devaluation after limited training. The administration of the dopamine antagonists α-flupenthixol and SCH23390 during training had no impact on this sensitivity; it was neither enhanced nor attenuated by these drugs. Eticlopride treatment, however, appeared to reduce sensitivity to the changed value of the reinforcer after taste aversion as evidenced by comparable rates of responding across the two devaluation groups. However, there was no statistical evidence for this effect, any inferences from Experiment 13 about the role of D_2 receptors in the control of goal-directed behaviour in normal animals would be premature.

Nevertheless, the results from the reinforcer devaluation task in animals with prior exposure to amphetamine (Experiments 10-12) furnish unequivocal evidence for distinct roles of D_1 and D_2 receptor subtypes in the control of behaviour by goal-directed actions and S-R habits. In a replication of the previous chapter's findings, animals given pre-training exposure to amphetamine but saline during training showed accelerated habit formation as they failed to alter lever press performance in response to the changed value of the reinforcer. The performance in the reinforcer devaluation task of sensitized animals treated with either the non-selective dopamine antagonist α-flupenthixol or the D_1 antagonist SCH23390 during training was not autonomous of the current value of the reinforcer as these animals showed a selective depression in lever press rates compared to

non-devalued controls. Thus the instrumental performance of these animals mirrors that of normal animals after limited training and suggests response control was by goal-directed A-O associations. The finding, however, that animals given eticlopride were impervious to the current value of the reinforcer suggests these animals' instrumental performance remained stimulus-bound and governed by S-R habits.

The differential sensitivity to outcome devaluation procedures cannot be attributed to impaired acquisition, as responding in all animals was depressed during acquisition irrespective of the antagonist administered. Similarly, all animals acquired an aversion to the reinforcer as evidenced by the marked sensitivity of magazine approach to outcome value in the extinction tests. Furthermore, the magazine entry data suggest that the eticlopride treated animals were able, under certain circumstances, to inhibit specific responses and hence the insensitivity of lever pressing in these animals cannot be accounted for by general response perseveration. Although the results of the activity test indicated elevated locomotor activity in eticlopride-treated animals in response to an amphetamine challenge compared to other animals, the sensitivity of magazine entry behaviour renders any interpretation of lever press performance in terms of hyperactivity most unlikely. The results are therefore specific to an effect on lever pressing and demonstrate that the accelerated habit formation following amphetamine exposure reported in Chapter 4 and replicated here, is prevented by D_1 but not D_2 receptor antagonism. Indeed, this parallels good evidence that the development of sensitization to the locomotor activating effects of amphetamine is also blocked by D_1 antagonists. These effects have been observed systemically (Vezina and Stewart, 1989) and with local

infusions of SCH23390 into both the VTA and substantia nigra pars reticulata (Stewart and Vezina, 1989; Vezina, 1996). Similarly, D_1 receptor knock-out mice fail to develop behavioural sensitivity to amphetamine (Karper et al., 2002; Mcdougall et al., 2005) and a recent fMRI study supports the suggestion that D_1 receptors are responsible for amphetamine-mediated neurochemical changes and that D_1 antagonists inhibit this response to amphetamine (Dixon et al., 2005). Thus the current findings concur with reports of D_1 receptor modulation of the neurochemical and locomotor response to amphetamine and extend them to include a further behavioural response; enhanced habit formation. More generally, they provide evidence of D_1 receptor involvement in the transition from actions to habits and raise the intriguing possibility that antagonism of D_1 receptors over the course of extended training would attenuate habit formation. Similarly, D_1 receptor knock-out mice may fail to develop goal-insensitive habitual responding.

However, eticlopride administered during training failed to reverse the accelerated formation of S-R habits induced by pre-training amphetamine exposure. This finding is consistent with growing evidence that D_2 antagonism actually enhances the behavioural and neurochemical effects of amphetamine. For example, the blockade of D_2 receptors in the VTA produces persistent elevation of the locomotor activating effects of amphetamine (Tanabe, Suto, Creekmore, Steinmiller and Vezina, 2004). Indeed, in the current experiments systemic administration of eticlopride during training appeared to heighten the potentiation of locomotor activity by amphetamine in both sensitized and non-sensitized animals in the activity test following a drug challenge. Sulpiride, which

has high affinity for D_2 receptors, has been shown to enhance the augmentative effects of amphetamine on extracellular striatal dopamine levels measured by in vivo microdialysis (Jaworski, Gonzales and Randall, 2001). Similarly, fMRI measurement of changes in rat brain activation following amphetamine administration shows that pre-treatment with sulpiride facilitates the response elicited by amphetamine (Dixon et al., 2005). Furthermore, the finding that the instrumental performance of animals treated with eticlopride was completely independent of goal-value during the reacquisition test also suggests that antagonism of D_2 receptors enhanced the effect of pre-training exposure to amphetamine on the sensitivity of a moderately trained instrumental response to outcome devaluation. The amphetamine-sensitized animals treated with eticlopride clearly had a representation of the devalued outcome as they inhibited magazine entry responses and when given the opportunity consumed less of the outcome compared to controls, but they failed to use this representation to guide instrumental responding. Instrumental performance under the control of S-R habits, whether engendered by overtraining or amphetamine exposure, is normally sensitive to outcome value in re-acquisition and thus the insensitivity of eticlopride-treated animals in this test in Experiment 12 is novel and can be taken as evidence of dysfunctional habit learning characteristic of compulsions. By definition, compulsive behaviour is carried out repetitively and persists despite adverse consequences. Significantly, there is emerging evidence that abnormal D_2 receptor binding may be involved in psychopathologies characterised by compulsive behaviour. For example, PET scans have revealed low D_2 receptor availability in drug abusers (Wang et al., 1997; Volkow et al., 1999; Volkow et al., 2001) and single photon emission computerized tomography (SPECT) has shown reduced D_2 receptor binding in

OCD patients (Denys, van der Wee, Janssen, De Geus and Westenberg, 2004). The current results are consistent with these reports and suggest that sensitization of dopaminergic systems coupled with the down-regulation of D_2 receptors may lead to maladaptive habitual behaviour that is compulsive. As such the paradigm developed here could serve as model of the neurochemical changes that accompany the loss of voluntary control over behaviour associated with drug addiction and neuropsychiatric disorders such as OCD and Tourette's Syndrome.

The finding of opposing roles of D_1 and D_2 receptors in the transition from action to habit and compulsion in the experiments presented here is consistent with previous reports that antagonism of D_1 receptors disrupts, but D_2 receptor blockade facilitates, learning in a variety of Pavlovian conditioning paradigms (Smith, Neill and Costall, 1997; Horvitz, 2001; Eyny and Horvitz, 2003; Yue et al., 2004, Cassaday, Nelson and Norman, 2005). The demonstration here of dissociable effects of D_1 and D_2 receptor antagonism on instrumental learning and the sensitivity of that learning to outcome devaluation is, however, novel. The effects of D_1 and D_2 receptors antagonists on learning mirror their effects on synaptic plasticity within the striatum with D_1 and D_2 receptors acting synergistically to allow LTD but in opposition during LTP (Kerr and Wickens, 2001; Centonze et al., 2001). This differential involvement in striatal synaptic plasticity may therefore underlie the effects on learning seen here and more generally accelerated habit formation after sensitization (Gerdeman, Partridge, Lupica and Lovinger, 2003).

CHAPTER 6

GENERAL DISCUSSION

Summary of findings

This thesis has presented a series of experiments that examined the behavioural and neurochemical processes involved in the development of S-R habits in rats. The results support previous findings for a dissociation of psychological and neural systems that control voluntary, goal-directed behaviours and reflexive, stimulus-bound habitual responding respectively (Adams, 1982; Dickinson et al., 1983; Dickinson, 1985; Dickinson and Balleine, 1993; Balleine and Dickinson, 1998; Killcross and Coutureau, 2003; Coutureau and Killcross, 2003; Yin et al., 2004; Yin et al., 2005). The current findings add considerably to our understanding of how these systems interact at a psychological and neurochemical level to control behaviour.

Behavioural analysis of the development of S-R habits in rats

The experiments in Chapters 2-3 provided a behavioural analysis of factors that contribute to the development of S-R habits. Significantly, these experiments replicated the ubiquitous finding that overtraining of an instrumental response engenders behaviour that is resistant to reinforcer devaluation. Although a pervasive view in Folk psychology, the results of these experiments demonstrate unambiguously that the development of habits is not simply the product of response repetition *per se*. Similarly, the finding in Chapter 3 that overtrained responses can still be guided by outcome expectancy suggests that extended exposure to the reinforcer is in itself not sufficient to generate habitual

responding. Furthermore, in contrast to previous research (e.g. Dickinson et al., 1983) these experiments found no evidence that the class of reinforcement schedule employed during training can influence the rate at which goal-independent responding develops. Firstly, Experiment 1 demonstrated that interval schedules do not necessarily facilitate habit formation. Indeed, animals in Experiment 1 showed sensitivity to outcome devaluation even after 240 response-reward pairings. Conversely, there was no evidence from Experiments 2 or 4 that training on ratio schedules attenuates the onset of goal-insensitive responding. As these two classes of reinforcement schedule establish markedly different behaviour-reward functions, the finding that habits developed at equivalent rates on these schedules suggests that the development of S-R habits may not simply be explained in terms of the relationship between behaviour and reward. Thus the current experiments provided no evidence to support previous suggestions that habits develop because overtraining an instrumental response (and training on interval schedules) degrades the relationship between behaviour and reward (Dickinson, 1985). Indeed, in Chapter 3 manipulations that affected this relationship rendered overtrained instrumental responses goal-directed irrespective of whether they introduced a stronger or weaker relationship between an animal's behaviour and reward delivery. Similarly, in Experiment 5 changes in reward magnitude did not affect the schedule of reinforcement *per se* but these manipulations nonetheless resulted in behaviour that was sensitive to outcome devaluation at test. The finding that animals could remain acutely sensitive to changes in the instrumental contingency after extended training is perhaps striking. It suggests that S-R habits do not develop as a result of changes in animals' knowledge of the A-O contingency, as otherwise performance would presumably be unaffected by the

behavioural manipulations introduced in Chapter 3. As these manipulations disrupted the normal overtraining effect of goal-insensitive performance, these results demonstrate unequivocally that A-O associations are not lost over the course of extended instrumental training and responding can rapidly come once again under the control of outcome expectancy. As such, these results are consistent with evidence from lesion studies that response control can switch between the habit and goal-directed systems (e.g. Coutureau and Killcross, 2003; Yin et al., 2004). Indeed, there is evidence from the medium training groups in Experiments 1 and 2 that A-O and S-R processes can be engaged concurrently. Responding in these animals appeared to be sensitive to both the selective and non-selective effects of pre-feeding to satiety. Thus, it would appear that the development of S-R habits is a dynamic process that involves competition between the goal-directed and habit systems: A-O associations are gradually overridden by inhibitory mechanisms that allow S-R processes to dominate behavioural expression.

Critically, the current experiments highlight factors that contribute to the relative dominance of one system over the other. In Chapter 2, detailed analysis of multifarious factors that could potentially account for animals' sensitivity to outcome devaluation was performed. This analysis revealed that there was no relationship between an animal's gross performance on the response-reward curve and sensitivity to outcome devaluation at test. There was no correlation between lever press rates or reward rates and test performance. Rather, the data suggested that S-R habits emerged as animals attained a consistent level of performance. On the interval schedule in Experiment 1, there was an inverse relationship between variation in response rates across the final session of

instrumental training and sensitivity to outcome devaluation at test. Animals that produced the least variation in response rates on the final day of acquisition showed the greatest resistance to reinforcer devaluation. In Experiment 2, habitual responding appeared to develop as animals pressed the lever at a consistent rate and no longer experienced the large session by session increases in rates typical of the early stages of acquisition on ratio schedules. Thus in broad terms, both these experiments revealed that instrumental performance comes to be controlled by goal-insensitive S-R habits as behaviour ceases to vary. This position was further supported by the results from the experiments presented in Chapter 3. These data demonstrated that once animals had established a stable rate of responding over 9 sessions of instrumental training but were then exposed to either changes in the rate of reinforcement or reward magnitude, instrumental performance in these animals appeared to be guided by outcome value. Significantly, the effect of all these manipulations was to lower response rates relative to baseline. This suggests that these manipulations may have rendered instrumental responding goal-directed as a direct consequence of their disruptive effects on well-established responding.

One interpretation of these findings is that habitual responding emerges as animals settle to a response rate that assures a predictable and stable relationship between their behaviour and reward delivery. This is perhaps most poignantly demonstrated by the finding from Experiment 6 that animals exposed to an ever-changing reinforcement schedule (and consequently were prevented from establishing a consistent relationship between behaviour and reward) failed to develop S-R habits despite extended training.

Once the relationship between behaviour and reward becomes well-predicted, animals need no longer attend to the consequences of their actions and hence instrumental performance becomes independent of outcome value. This suggests that S-R habits are not the product of changes in animals' knowledge of the relationship between behaviour and reward but rather goal-insensitive responding emerges as this relationship becomes well-predicted. This position stands in stark contrast to previous theoretical accounts that postulated that S-R habits arose due to a decline in the experienced correlation between behaviour and reward (e.g. Dickinson, 1985). Furthermore, the finding that habits developed as animals obtained stable response rates mirrors findings from neurophysiological studies that over the course of procedural learning neuronal responses within the sensorimotor striatum undergo a process of restructuring and fine-tuning. As performance becomes well-established, neuronal activity shifts from individual tasks elements to marking only the beginning and end of the task (Jog et al., 1999; Barnes et al., 2005). This is consistent with suggestions that the sensorimotor striatum builds up or 'chunks' elements of actions into a template of the entire procedure (Graybiel, 1998; Graybiel, 2004). Seen in this context, the current findings are also consistent with more generic examples of habit learning: complex sequences of behaviour, such as driving, come to be performed effortlessly as single units triggered by specific environmental stimuli.

Neurochemical analysis of the development of S-R habits in rats

A further significant contribution of the experiments reported here is greater elucidation of the neurochemical processes involved in the transition from goal-directed to habitual

stimulus-bound forms of behaviour. Dopamine-dependent synaptic plasticity has been widely implicated in reward-related learning (e.g. Berke and Hyman, 2000; Reynolds et al., 2001; Schultz, 2002; Wickens et al., 2003). However, relatively few studies have examined its role in S-R habit learning (e.g. Robbins et al., 1991; Packard and White, 1991). Furthermore, as none of these studies explicitly probed the associative structure underpinning performance, it cannot be directly inferred that these tasks involved S-R habit learning. However, the application in the current studies of behavioural assays, such as reinforcer devaluation, provides rather better evidence of the extent to which behaviour is controlled by voluntary goal-directed actions or reflexive habits.

In Experiment 8, animals were exposed for 7 days to the indirect dopamine agonist amphetamine prior to the commencement of instrumental training. Although the instrumental response was only moderately trained (3 sessions), animals' performance in two subsequent tests of goal-sensitivity was found to be independent of the current value of the reward. However in Experiment 9, responding in animals that received equivalent exposure to amphetamine after the completion of instrumental testing but before extinction testing, was shown to be guided by outcome expectancy. This dissociation between pre- and post-training amphetamine exposure is significant because it suggests that the failure to detect normal goal-sensitive responding after moderate training in Experiment 8 cannot be accounted for by any effects of amphetamine on motivation or locomotor activity, as presumably such effects would have also be observed in Experiment 9. Furthermore, evidence that magazine entry behaviour remained sensitive to outcome devaluation, coupled with the results from the reacquisition test, demonstrates

that the insensitivity of lever press performance in the extinction test to reinforcer devaluation did not result from a failure to acquire a representation of the changed value of the reward. Rather, it suggests that animals were unable to use this representation to guide instrumental performance. Thus, these findings mimic both the effect of overtraining seen in normal animals as well as the effects produced by lesions to the prelimbic cortex and the dorsomedial striatum (Killcross and Coutureau, 2003; Yin et al., 2005a). As the control of instrumental performance has been posited to be determined by active competition between the goal-directed and habit systems, the effect of amphetamine sensitization seen here could be mediated by two possible processes. Firstly, amphetamine exposure may have facilitated the development of S-R habit learning *per se* or alternatively, it may have disrupted A-O processes and allowed *developing* S-R habits to dominate instrumental performance prematurely. The finding that amphetamine sensitization affected the acquisition, but not the expression, of instrumental learning clearly favours the former hypothesis. Thus the demonstration of accelerated habit-formation after sensitization with amphetamine suggests a fundamental role of dopamine in the 'stamping in' of S-R associations that underpin procedural learning.

These effects were further explored in Chapter 5. Specifically, these experiments sought to expound the neuropharmacological selectivity of the sensitization effect observed in Chapter 4. Animals were firstly sensitized to amphetamine and then, as in Chapter 4, received 3 sessions of instrumental training. Prior to each of these sessions, animals were administered either selective or non-selective dopamine antagonists. Finally, in an

attempt to explore more generally the role of dopamine in instrumental learning, non-sensitized animals were administered dopamine antagonists during the acquisition of a moderately trained response.

The general effect of antagonism of dopaminergic systems during training was to suppress instrumental performance. Interestingly, this effect was seen whether or not animals had received prior exposure to amphetamine. This finding is consistent with previous reports that dopamine antagonists produce decreases in response rates (e.g. Varvel et al., 2002). However, the current experiments suggested that these effects were not uniform and depended on the selectivity of the drug. As the effects of both α-flupenthixol and the selective D_2 antagonist eticlopride on responses rates were restricted to the training phase and were not present in the drug-free tests, it suggests that these drugs did not affect the rate of learning, but rather that their depressive effects on instrumental performance were mediated by motivational, attentional or locomotor processes. On the other hand, animals administered the selective D_1 antagonist SCH23390 during training also displayed lower response rates compared to saline controls in the drug-free tests conducted 4 and 5 days after the final injection of SCH22390. In view of this time frame, the reduced rates of performance seen in these tests can not be attributed simply to non-associative factors such as motivation and locomotor activity. These results suggest, therefore, that antagonism of D_1 receptors by SCH23390 affected not only the performance of the instrumental response but actually attenuated the rate or consolidation of associative learning. This finding is consistent with evidence that stimulation of D_1 receptors contributes to dopamine-dependent synaptic

plasticity that has been postulated to be necessary for learning (Schultz, 2002; Wickens et al., 2003). For example, activation of dopamine D_1 receptors by either endogenous dopamine or exogenous dopamine agonists has been shown to induce long-term potentiation within the striatum (Kerr and Wickens, 2001). Similarly, electrical stimulation of the substantia nigra induces potentiation of corticostriatal synapses but this effect is blocked by SCH23390 (Reynolds et al., 2001).

This interpretation is further supported by the finding that amphetamine-induced enhancement of S-R habits was reversed by both the non-selective dopamine antagonist α-flupenthixol and the selective D_1 antagonist SCH23390. At test, animals exposed to amphetamine and then administered these compounds during training showed normal sensitivity to outcome devaluation despite pre-training exposure to amphetamine. This suggests that accelerated habit-formation seen after amphetamine sensitization is in part mediated by activation of D_1 receptors. This is consistent with previous reports indicating that the behavioural and neurochemical response to psychostimulants such as amphetamine is modulated via D_1 receptors (e.g. Xu et al., 1994; Xu, Guo, Vorhees and Zhang, 2000; Zhang et al., 2005). A markedly different pattern of results emerged in amphetamine sensitized-animals treated with the D_2 antagonist eticlopride. These animals showed the same abnormal insensitivity to outcome devaluation after moderate training as control animals. Thus in stark contrast to the effects of SCH23390, antagonism of D_2 receptors failed to reverse amphetamine induced disruption of goal-directed responding. Furthermore, in the subsequent reacquisition test lever press performance in these animals appeared to remain completely impervious to outcome value. Evidence from

both magazine approach behaviour and the consumption test confirmed that these animals had acquired an aversion to the reinforcer. This suggests that these animals were completely impaired in their ability to use the current value of the reward to guide instrumental performance, even when presented with the consequences of their actions in the reacquisition test. Normally, when instrumental responses under the control of S-R habits and are rewarded with an outcome that has been paired with gastric-malaise, animals' propensity to press the lever is markedly reduced. Thus, sensitization with amphetamine coupled with blockade of D_2 receptors not only led to the early and excessive dominance of S-R processes, it appeared to render instrumental performance in these animals compulsive. There is increasing evidence of abnormal D_2 receptor binding in OCD patients and compulsive drug users (e.g. Denys et al., 2004; Volkow, Fowler, Wang and Swanson, 2004; Nader and Czoty, 2005; Nader et al., 2006). As such, the current findings have implications for understanding the neuropharmacological mechanisms involved in the development of abnormal habitual behaviour associated with certain human psychopathologies. More generally, these findings demonstrate distinct contributions of D_1 and D_2 receptors to the normal development of S-R habits. This development may involve the activation of D_1 receptors combined with reduced transmission at D_2 receptors. The current findings are consistent with previous reports of differential involvement of these receptors in learning (e.g. Horvitz, 2001) and it is suggested that this may reflect distinct contributions of these receptors to striatal synaptic plasticity (Kerr and Wickens, 2001; Centonze et al., 2001).

Optimality, uncertainty and dopamine - the development of S-R habits

The current research has highlighted both behavioural and neurochemical processes involved in the transition of behavioural control over the course of instrumental training.

From a behavioural perspective, the current experiments demonstrated that the development of S-R habits was associated with the achievement of stable rates of performance. It is of central importance to establish how and why this process of stabilization occurs. Some of the most compelling and parsimonious accounts of instrumental performance are those that assume that all behaviour constitutes choice and that choice is driven by feedback from the interaction between behaviour and the environment (e.g. Herrnstein, 1970; Baum, 1973). Even in the restricted environment established by the experimenter, it is clear that an animal always has more than one activity in which it can engage. For example, apart from measured behaviour such as lever pressing, animals can explore, groom or rest. Engaging in one activity inextricably leads to a decline in the performance of other behaviours. It is clear that choosing between different behaviours produces not only different rewarding consequences but also different costs, characterised as different requirements for effort. Optimization accounts of instrumental performance assume that the goal of behaviour is to maximize satisfaction and minimize effort. In order to optimize satisfaction, animals must trade-off the costs and benefits associated with any given behaviour (e.g. Baum, 1981). There is very good evidence to suggest that animals' behaviour is determined by some form of cost-benefit analysis (e.g. Salamone et al., 1991; Salamone, Cousins and Bucher, 1994;

Cousins, Atherton, Turner and Salamone, 1996; Walton, Bannerman, Rushworth, 2002). These studies have shown that animals are sensitive to changes in the cost (e.g. the effort required to obtain a reward) and the benefit (e.g. reward magnitude) of a given activity. When the cost of an activity is increased without a concomitant increase in benefit, animals tend to shift behaviour towards activities associated with fewer rewards but also less effort (e.g. van den Bos, van der Harst, Jonkman, Schilders and Sprijt, 2006). Similarly, the finding that ratio schedules maintain higher response rates than interval schedules can be readily explained by optimality theories of instrumental performance (Baum, 1993). Of course, on any class of reinforcement schedule higher rates lead to greater expenditure of energy and fewer opportunities to pursue other activities. If it is assumed that this positively accelerating cost function is equivalent on interval and ratio schedules, it follows that rates will be lower on an interval compared to a ratio schedule. As interval schedules impose an upper limit on the rate of reinforcement, increasing the response rate beyond a certain level will only entail an increase in cost. On the other hand, on a ratio schedule higher rates are associated not only with increased cost but also a more rewarding reinforcement rate. Thus the potential for net gain from the trade-off between cost and benefit is greater on a ratio schedule. This relationship is depicted in Figure 6.1.1. To test informally these predictions, rats were trained in two identical boxes and received one session in each box for 15 days. In one box animals were trained on an RI30 schedule and in the other on a RR8 schedule. As the boxes were identical and animals were rewarded with the same reinforcer in each box (food pellets), the costs associated with each response should have been equivalent. Similarly, by comparing performance on the two schedules in the same animal any effects due to individual

differences could be obviated. After 15 days of training, animals were responding approximately 20 times a minute on the interval schedule and 30 times on the ratio schedule.

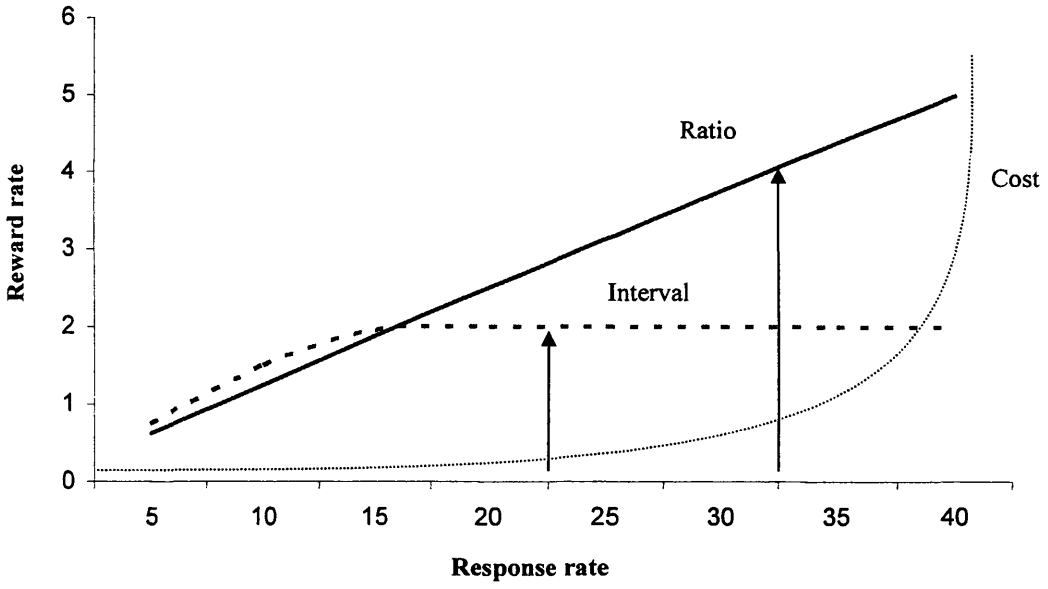

Figure 6.1.1 Cost and benefit function of interval (dashed black lines) and ratio schedules (solid black lines). Arrows depict mean response rate after 15 sessions of interval (RI30) and ratio training (RR8). Grey lines represent the accelerating cost function. (Adapted from Baum, 1981).

It is clear from the portrayal of these data in Figure 6.1.1 that animals settled for rates that assured a near optimal trade-off between cost and benefit: the increase in cost associated with the higher rates of performance on the ratio schedule was offset by the increase in the rate of reward. Similarly, pressing the lever at a higher rate on either schedule would

have brought diminishing returns as assessed by the relative rate of increase of cost and reward functions. These data are consistent with the suggestion that instrumental performance is determined to some extent by a trade-off between cost and benefit. Seen in this context, the finding in Chapters 2 and 3 that S-R habits developed as animals attain a stable rate of performance is significant. It not only explains why response rates stabilize, it also suggests that goal-insensitive habitual responding may be partly a product of the trade-off between cost and benefit. If instrumental performance is driven by the goal of maximizing benefits and minimizing costs, during the early stages of instrumental learning animals must explore the relationship between their behaviour and reward delivery in order to estimate the long-term value of a given rate of responding. Clearly, such a cost-benefit analysis places considerable demands on cognitive resources such as working memory and attention. It requires both a representation of the outcome and the nature of the causal relationship between pressing the lever and reward delivery (i.e. the instrumental contingency). These are, of course, the cardinal features of goal-directed actions. Changing the value of the reinforcer during the initial stages of training, while animals are still ascertaining the level of performance necessary to maximize gain and minimize cost, would clearly impact on their behaviour. However, once animals have settled for a rate of performance that secures an optimal or near optimal trade-off between cost and benefit, parsimony dictates that attending to and encoding outcome value would become superfluous (cf. Pearce and Hall, 1980). Hence once animals reach stable rates of performance, responding is no longer guided by outcome-value. Of course, this only holds if the relationship between reward and behaviour is kept constant. Introducing variation in these factors would render instrumental performance uncertain and

potentially suboptimal. Thus in order to optimise long-term behaviour, animals would have to perform a new cost benefit-analysis to reflect the now changed environmental contingencies. This could account for the restoration of goal-directed responding observed in animals that experienced sudden variation in both the rate of reinforcement and reward magnitude in Experiments 3-5. Similarly, an ever-changing environment would prevent animals from establishing an optimal level of trade-off. This was evidenced in Experiment 6 as animals that were exposed to an inconsistent reinforcement rate failed to develop habitual responding.

Significantly, effort-based cost-benefit evaluations in free-operant tasks have been shown to be highly sensitive to dopaminergic manipulations and in particular to depletion of accumbens dopamine (e.g. Aberman, Ward and Salamone, 1998; Aberman and Salamone, 1999; Salamone, Wisniecki, Carlson and Correa, 2001; Mingote, Weber, Ishiwari, Correa and Salamone, 2005). In general, these studies have shown that either the administration of dopamine antagonists or local infusions of 6-OHDA into the nucleus accumbens reduces the effort animals are willing to expend to obtain rewards as evidenced by increased ratio strain but is without effect on the consumption of freely available rewards. Taken together these data suggest that dopamine may be involved in the process whereby animals evaluate the cost and benefits of an action. Indeed, a recent theoretical paper has applied reinforcement learning models in an attempt to explain how dopaminergic tone may affect response vigour (Niv, Daw, Joel and Dayan, 2006). It is argued that the average reward rate is critical to determining the cost-benefit trade-offs that allow animals to evaluate how much effort to expend to obtain available rewards.

Furthermore, the authors suggest that the average reward rate is reported by tonic dopamine. This would account for the above mentioned findings that dopamine depletions can reduce the effort animals are prepared to expend in order to obtain rewards. Although this theory may in part explain the reduced response rates seen in Chapter 5 after pre-treatment with dopamine antagonists, it is unlikely that it can account for the enhancement of S-R habit learning seen after amphetamine sensitization. For example, it has been shown that hyperdopaminergic mice with a 70% higher level of tonic dopamine show elevated response rates compared to wild-types but normal A-O learning after limited training, as indexed by sensitivity to outcome devaluation (Yin, Zhuang and Balleine, 2006b). This suggests that increased dopaminergic tone does not necessarily lead to accelerated habit formation. However, as pharmacological manipulations of dopamine are likely to affect both tonic and phasic signalling, the effects seen in Chapters 4 and 5 could also be mediated by changes in phasic dopamine activity.

In broader terms, the effect of stable and near-optimal levels of performance is that the long-term molar correlation between behaviour and reward becomes well-predicted. This is consistent with the neurocomputational model of these processes outlined in Chapter 1 (Daw et al., 2005). These authors propose a dual-action choice system based on two classes of reinforcement learning. They postulate that goal-directed behaviour is subserved by 'model-based' methods or 'tree search'. This involves constructing predictions of long-run outcomes by exploring the molecular correlation between behaviour and reward and hence predictions are based on the immediate consequences of

actions. On the other hand, response control by the habit system is argued to be founded on 'caching' which refers to the association of an action with a scalar summary of its long-term value. Such a method does not require animals to explore exhaustively the relationship between behaviour and reward, but rather action is based on an estimation of long-term value. 'Tree search' models concur with the suggestion above that during the early stages of learning animals explore the correlation between behaviour and reward in order to secure an optimal trade-off between cost and benefit. Similarly, the finding in Chapter 2 that habits developed as animals established stable rates of responding is consistent with a 'caching' model of habitual response control. Stable response rates imply that the relationship between behaviour and the occurrence of rewards has become well-predicted and that animals have learnt that a certain response rate is associated with a long-term estimation of value. Furthermore, as outlined in the introduction, Daw et al. propose that competition between the two systems is mediated by uncertainty, such that responding is controlled by the more accurate of the two systems. This would account for the results found in Chapter 3 that disrupting well-established performance restores response-control to the goal-directed system. Unexpected changes in task contingencies clearly favours response-control based on 'tree search' methods. 'Tree search' provides a more efficient method of estimating the value of actions in changed circumstance compared to inflexible cached values.

The suggestion that habits develop as the molar relationship between behaviour and reward becomes well-predicted is consistent with contemporary learning theories that attach critical importance to the role of reward predictability in driving learning. Indeed,

the reward prediction error, or the discrepancy between expected and actual rewards, is a central tenet of modern learning theories (e.g. Rescorla and Wagner, 1972; Mackintosh, 1975; Pearce and Hall, 1980). According to these theories, learning only proceeds when the occurrence of rewards is better or worse than predicted: fully predicted rewards do not contribute to learning. Significantly, there is very good evidence that the activity of midbrain dopamine neurons provides a neurophysiological correlate of the reward prediction error signal predicated in contemporary learning theories (e.g. Schulz, 1998; Waelti et al., 2001). These studies have shown that the firing rate of midbrain dopamine neurons is unaffected by rewards that are as good as predicted but increase in response to unexpected rewards and is suppressed by the omission of expected rewards (Ljungberg et al., 1992; Schultz, Apicella and Ljungberg, 1993; Mirenowicz and Schultz, 1994). Furthermore, it has been argued that the dopamine prediction error may serve as a 'teaching signal' that can modulate cortico-striatal synaptic plasticity that underpins reward-driven learning (Schultz, 1998; Schultz, 2002; Schultz, Tremblay and Hollerman, 2003). It is clear that such a system allows organisms to model the causal relationship of its interactions with the environment and consequently adapt its behaviour to optimize long-term gain. Significantly, it has been shown that over the course of training the phasic dopamine response is transferred from rewards to the stimuli that predict the occurrence of rewards. (Schultz et al., 1997). This effect may represent a neurophysiological analogue of the transfer of response control over the course of extended instrumental training from its consequences to its antecedents. The findings from Chapter 2 that instrumental performance became goal-insensitive as animals established stable response rates supports the suggestion that as reward delivery becomes

well-predicted animals no longer attend to the consequences of their actions. Although the vast majority of the work implicating phasic dopamine in encoding a reward prediction error has been conducted using appetitive Pavlovian procedures, more recent work has also demonstrated a role for the phasic dopamine prediction error signal in instrumental learning (e.g. Morris, Arkadir, Nevet, Vaadia and Bergman, 2004; Bayer and Glimcher, 2005; Morris, Arkadir, Nevet, Vaadia and Bergman 2006; Pessiglione, Seymour, Flandin, Dolan and Frith, 2006). Interestingly, one of these studies has shown that even after animals have developed a fixed behavioural response, midbrain dopamine neurons continue to compute a reward prediction error but animals' behaviour is largely unaffected by the error term (Bayer and Glimcher, 2005). This may account for the finding in Chapter 2 that variation in local reward rates did not correlate with animals' sensitivity to outcome devaluation and is consistent with the suggestion that habits develop as the molar, rather than the molecular, correlation between reward and behaviour becomes well-predicted. Indeed, in Chapter 3 it was shown that when this relationship is sufficiently disrupted response control is restored to the goal-directed system. Thus, the finding that midbrain dopamine neurons continue to compute a prediction error even after extended training may provide a mechanism for the effects observed in Chapter 3.

Similarly, the effects reported in Chapters 4 and 5 may also be attributable to the phasic activity of dopamine neurons. As the blockade of dopamine reuptake by amphetamine leads to sustained increases in extracellular dopamine levels, the effect of the psychostimulant drugs would be to exaggerate the phasic response of dopamine neurons

to rewards and reward-related stimuli (Schultz, 2002). For example, the effect of an exaggerated reinforcement signal may be to accelerate the transfer of dopamine neuronal firing from rewards to reward-predicting stimuli. The magnified 'teaching signal' encoded by dopamine neurons would also exert considerable influence over cortico-striatal synaptic plasticity that has been proposed to modulate learning. Similarly through its effects on tonic extracellular concentrations of dopamine, amphetamine may also amplify the responses of dopamine neurons in target structures, such as the prefrontal cortex and striatum that receive excitatory input from the midbrain dopamine prediction error signal. This may lead to the early and preferential recruitment of the neural circuitry outlined in Chapter 1 that subserves response control by S-R habits. Taken to together, these effects may serve to enhance associative learning and hence underpin the facilitation of S-R learning seen after amphetamine sensitization in Chapters 4 and 5 (Berke and Hyman, 2000).

Wider implications and future directions

The experiments reported here add significantly to our understanding of both the psychological and neurochemical processes involved in the transition from goal-directed to habit-based responding. Changes in the balance of these processes may be associated with certain human psychopathologies. For example, the demonstration here that psychostimulant drugs can hijack the habit system is highly significant for our understanding of the transition from drug abuse to compulsive drug-taking. Indeed, it is becoming increasingly recognised that drug addiction may in part be the product of aberrant habit learning (e.g. Everitt and Robbins, 2005; Nelson and Killcross, 2006;

Fuchs, Branham and See, 2006). The experiments reported in this thesis furnish significant new insights into the putative neurochemical processes involved in drug addiction and as such may provide a model of the development of habitual and eventually compulsive drug-taking. Similarly, as disruption to the voluntary control of behaviour is the hallmark of other neuropsychiatric conditions such as obsessive compulsive disorder and Tourette's syndrome, the current findings may have implications for our understanding of the psychopathogenesis of these conditions.

In this respect, it will be of central importance to elucidate the neuroanatomical structures that mediate the enhancement of S-R habit learning seen in Chapter 4 after sensitization with amphetamine. This could be achieved by combining sensitization with the reversible inactivation of structures implicated in the control of habitual behaviour. For example, if inactivation of the infralimbic cortex or the dorsolateral striatum in sensitized animals restores goal-directed responding then this would indicate that sensitization leads to the early and excessive control of the habit system over behaviour. Similarly, the distinct contributions of D_1 and D_2 receptors to these processes require further investigation. For instance, the results of Chapter 5 raise the possibility that animals treated with a D_1 antagonist over an extended period of training may fail to show the normal overtraining effect of goal-insensitive responding. Conversely, D_2 receptor antagonists administered for an equivalent period might render instrumental performance completely impervious to reward value. As the finding that sensitization coupled with D_2 down-regulation promotes compulsive lever pressing may represent a putative model of aberrant habit learning seen in drug addiction and other human psychopathologies, it will

be important to test the model's predictive validity by attempting to restore normal goal-directed responding through the administration of therapeutic compounds used in the treatment of these disorders (e.g. novel dopaminergic compounds and GABA agonists) (Sofuoglu and Kosten, 2005).

Furthermore in accordance with previous suggestions, the experiments reported here in Chapter 3 indicated that response control is determined to some extent by competition between the goal-directed and habit systems. To date the infralimbic cortex has been identified as a critical structure in regulating competition between these two systems (Killcross and Coutureau, 2003). However, there is also good evidence that the anterior cingulate is involved in response competition (e.g. Haddon and Killcross, 2006; de Wit, Balleine and Dickinson, 2006). Significantly, a recent paper has suggested that this structure is also important for learning the value of actions (Kennerley, Walton, Behrens, Buckley and Rushworth, 2006). As such, lesions to this structure may not affect the normal overtraining effect of goal-insensitive responding but might disrupt the restoration of goal-directed performance seen in Chapter 3 after animals experience changes in the rate of reinforcement or reward magnitude.

REFERENCES

Aberman, J. E., Ward, S. J. & Salamone, J. D. (1998). Effects of dopamine antagonists and accumbens dopamine depletions on time-constrained progressive-ratio performance. *Pharmacology, Biochemistry and Behavior,* 61(4), 341-348.

Aberman, J. E., & Salamone, J. D. (1999). Nucleus accumbens dopamine depletions make rats more sensitive to high ratio requirements but do not impair primary food reinforcement. *Neuroscience,* 92(2), 545-552.

Adams, C. D. (1980). Post-conditioning devaluation of an instrumental reinforcer has no effect on extinction performance. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology,* 32, 447-458.

Adams, C. D., & Dickinson, A. (1981a). Actions and habits: Variations in associative representations during instrumental learning. In Miller, R.R., & Spear, N.E. (Eds.), *Memory Mechanisms in Animal Behavior.* Hillsdale, N. J.: Erlbaum.

Adams, C. D., & Dickinson, A. (1981b). Instrumental responding following reinforcer devaluation. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 33B*(2), 109-121.

Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology,* 34B(2), 77-98.

Albin, R. L., Young, A. B., and Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends in Neuroscience,* 12, 366-375.

Alexander, G. E., Delong, M. R., & Strick, P.L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience,* 9, 357-381.

Allison, J. (1980). Conservation, matching and the vaiable-interval-schedule. *Animal Learning and Behavior,* 8, 185-192.

Aosaki. T., Tsubokawa, H., Ishida, A., Watanabe, K., Graybiel, A. M., & Kimura, M. (1994a). Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *Journal of Neuroscience,* 14(6), 3969-3984.

Aosaki. T., Graybiel, A. M., & Kimura, M. (1994b). Effect of the nigrostriatal dopamine system on acquired neural responses in the striatum of behaving monkeys. *Science,* 265(5170), 412-415.

Balleine, B. W. (1992). Instrumental performance following a shift in primary motivation depends on incentive learning. *Journal of Experimental Psychology: Animal Behavior Processes,* 18(3), 236-250.

Balleine, B. W., Garner, C., Gonzalez, F., & Dickinson, A. (1995). Motivational control of heterogenous instrumental chains. *Journal of Experimental Psychology: Animal Behavior Processes, 21,* 203-217.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology, 37*(4-5), 407-419.

Balleine, B. W., Killcross, A. S., & Dickinson, A. (2003). The effect of lesions of the basolateral amygdala on instrumental conditioning. *Journal of Neuroscience, 23*(2), 666-675.

Barnes, T. D., Kubota, Y., Hu, D., Jin, D., & Graybiel, A.M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding or procedural memories. *Nature,* 437, 1158-1161.

Barrot, M., Marinelli, M., Abrous, D. N., Rouge-Pont, F., Le Moal, M., & Piazza, P. V. (1999). Functional heterogeneity in dopamine release and in the expression of Fos-

like proteins within the rat striatal complex. *European Journal of Neuroscience, 11*(4), 1155-1166.

Baum, W. M. (1973). The correlational-based law of effect. *Journal of the Experimental Analysis of Behaviour,* 20, 137-143.

Baum, W. M. (1981). Optimization and the matching law as accounts of instrumental behavior. *Journal of the Experimental Analysis of Behaviour,* 36(3), 387-403. *Journal of the Experimental Analysis of Behaviour*

Baum, W. M. (1993). Performances on ratio and interval schedules of reinforcement: Data and theory. *Journal of the Experimental Analysis of Behaviour,* 59(2), 245-264.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron,* 47, 129-141.

Bayer, S. A. (1990). Neurogenetic patterns in the medial limbic cortex of the rat related to anatomical connections with the thalamus and striatum. *Experimental Neurology*, 107(2), 132–142.

Bechara, A., Dolan, S., Denburg, N., Hindes, A., Anderson, S.W., & Nathan, P.E. (2001). Decision-making deficits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in alcohol and stimulant abusers. *Neuropsychologia,* 39, 376–389.

Bentivoglio, M., & Morelli, M. (2005). The organisation and circuits of mesenchepahlic dopaminergic neurons and the distribution of dopamine receptors in the brain. In Dunnett, S.B., Bentivoglio, M., Björklund, A., & Hökfelt, T. (Eds.) *Handbook of Chemical Neuroanatomy, Volume 21: Dopamine.* Amsterdam: Elsevier.

Benninger, R.J., & Miller, R. (1998). Doapmine D1-like receptors and reward related incentive learning. *Neuroscience & Biobehavioral Reviews,* 22(2), 335-345.

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

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research Reviews,* 28(3), 309-369.

Blundell, P., Hall, G., & Killcross, S. (2001). Lesions of the basolateral amygdala disrupt selective aspects of reinforcer representation in rats. *Journal of Neuroscience,* 21, 9018-9026.

Bolles, R. C. (1972). Reinforcement, expectancy and learning. *Psychological Review,* 79, 394-409.

Boyson, S.J., McGonigle, P., & Molinoff, P.B. (1986). Quantitative autoradiographic localization of the D1 and D2 subtypes of dopamine receptors in rat brain. *Journal of Neuroscience,* 6(11), 3177-88.

Calabresi, P., Saiardi, A., Pisani, A., Baik, J.H., Centonze, D., Mercuri, N.B., Bernardi, G., & Borrelli, E. (1997). Abnormal synaptic plasticity in the striatum of mice lacking dopamine D2 receptors. *Journal of Neuroscience*, 17, 4536-4544

Canales, J. J. (2005) Stimulant-induced adaptations in neostriatal matrix and striosome systems: transiting from instrumental responding to habitual behavior in drug addiction. *Neurobiology of Learning and Memory, 83*, 93-103.

Canales, J.J., & Graybiel, A.M. (2000). A measure of striatal function predicts motor stereotypy. *Nature Neuroscience,* 3, 377-383.

Canales, J. J., Capper-Loup, C., Hu, D., Choe, E. S., Upadhyay, U., & Graybiel, A. M. (2002) Shifts in striatal responsivity evoked by chronic stimulation of dopamine and glutamate systems, *Brain, 125*, 2353–2363.

Catania, A.C., Matthews, T.J., Silverman, P.J., & Yohalem, R. (1977). Yoked variable-ratio and variable-interval responding in pigeons. *Journal of the Experimental Analysis of Behaviour,* 28(2), 155-161.

Capper-Loup C., Canales, J.J., Kadaba, N., & Graybiel, A.M. (2002). Concurrent activation of dopamine D_1 and D_2 receptors is required to evoke neural and behavioural phenotypes of cocaine sensitization. *Journal of Neuroscience,* 22, 6218-6227.

Carelli, R M., Wolske, M., & West, M.O. (1999). Loss of lever press-related firing of rat striatal forelimb neurons after repeated sessions in a lever pressing task. *Journal of Neuroscience,* 17, 1804-1814.

Cassaday, H.J., Nelson, A.J.D., & Norman, C. (2005). Haloperidol can increase responding to both discrete and contextual cues in trace conditioned rats. *Behavioural Brain Research*, 158(1), 31-42.

Centonze, D., Picconi, B., Gubellini, P., Bernardi, G., Calabresi, P. (2001). Dopaminergic control of synaptic plasticity in the dorsal striatum. *European Journal of Neuroscience*, 13, 1071-1077.

Childress, A.R., Mozley, P.D., McElgin, W., Fitzgerald, J., Reivich, M., & O'Brien, C.P. (1999). Limbic activation during cue-induced cocaine craving. *The American Journal of Psychiatry*,156(1), 11-18.

Choi, W.Y., Balsam, P.D., & Horvitz, J.C. (2005). Extended habit training reduces dopamine mediation of appetitive response expression. *Journal of Neuroscience,* 25(29), 6729-6733.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern analysing skills in amnesia: dissociation of knowing how and knowing that. *Science,* 210, 207-209.

Cole, M. R. (1994). Response-rate differences in variable-interval and variable-ratio schedules: An old problem revisited. *Journal of the Experimental Analysis of Behavior,* 61, 441–451.

Colwill, R. M., & Rescorla, R. A. (1985). Instrumental responding remains sensitive to reinforcer devaluation after extensive training. *Journal of Experimental Psychology: Animal Behavior Processes, 11*(4), 520-536.

Colwill, R. M., & Rescorla, R. A. (1988). The role of response-reinforcer associations increases throughout extended instrumental training. *Animal Learning & Behavior* , 11, 105-111.

Colwill, R. M., & Triola, S. M. (2002). Instrumental responding remains under the control of the consequent outcome after extended training. *Behavioural Processes, 57*(1), 51-64.

Corbit, L. H., Muir, J. L., & Balleine, B. W. (2001). The role of the nucleus accumbens in instrumental conditioning: Evidence of a functional dissociation between accumbens core and shell. *Journal of Neuroscience, 21*(9), 3251-3260.

Corbit, L. H., Muir, J. L., & Balleine, B. W. (2003). Lesions of mediodorsal thalamus and anterior thalamic nuclei produce dissociable effects on instrumental conditioning in rats. *European Journal of Neuroscience, 18*(5), 1286-1294.

Corbit, L. H.,& Balleine, B. W. (2003). Instruemental and Pavlovain incentive processes have dissociable effects on components of a heterogeneous instrumental chain. *Journal of Experimental Psychology: Animal Behavior Processes,* 29(2), 99-106.

Cousins, M.S., Atherton, A., Turner, L., & Salamone, J.D. (1996). Nucleus accumbens dopamine depletion alters relative response allocation in a T-maze cost/benefit task. *Behavioural Brain Research,* 74, 189-197.

Coutureau, E., & Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behavioural Brain Research, 146*(1-2), 167-174.

Davis, J., & Bitterman, M. E. (1971) Differential reinforcement of other behavior (DRO): a yoked-control comparison. *Journal of Experimental Analysis of Behavior,* 15, 237-241.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Actions, policies, values and the basal ganglia. In Bezard, E. (Ed.), *Recent Breakthroughs in Basal Ganglia Research.* Nova Science.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience,* 8 (12), 1704-1711.

Dawson, G. R., & Dickinson, A. (1990). Performance on ratio schedules and interval schedules with matched reinforcement rates. *Journal of Experimental Psychology: Comparative and Physiological Psychology,* 42B, (3), 225-239.

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

Denys, D., van der Wee, N., Janssen, J., De Geus, F., & Westenberg, H.G. (2004). Low level of dopaminergic D2 receptor binding in obsessive-compulsive disorder. *Biological Psychiatry,* 55(10), 1041-1045.

Devan, B. D., Mcdonald, R. J., & White, N. M. (1999). Effects of medial and lateral caudaute-putamen lesions on place- and cue-guided behaviour in the water maze: relation to thigmotaxis. *Behavioural Brain Research,* 100, 5-14.

Devan, B. D., & White, N. M. (1999). Parallel information processing in the dorsal striatum: relation to hippocampal function. *Journal of Neuroscience,* 19, 2789-2798.

De Wit, S., Kosaki, Y., Balleine, B.W., Dickinson, A. (2006). Dorsomedial prefrontal cortex resolves response conflict in rats. *Journal of Neuroscience,* 26(19), 5224-5229.

Di Chiara, G. D., & Imperato, A. (1988). Drugs Abused by Humans Preferentially Increase Synaptic Dopamine Concentrations in the Mesolimbic System of Freely Moving Rats. *Proceedings of the National Academy of Sciences,* 85(14), 5274-5278.

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society,* 308, 67-78.

Dickinson, A., Nicholas, D. J., & Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 35*(B-1), 35-51.

Dickinson, A., & Balleine, B. (1993). Actions and responses: The dual psychology of behaviour. In Eilan, B., McCarthy, R., & Brewer, B. (Eds), *Spatial Representations: problems in philosophy and psychology.* Oxford (England); Cambridge (Mass).: Blackwell.

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning and Behavior,* 22, 1-18.

Dickinson, A., Balleine, B., Watt, A., & Gonzalez, F. (1995). Motivational control after extended instrumental training. *Animal Learning and Behavior,* 23(2), 197-206.

Dickinson, A., Campos, J., Varga, Z. I., & Balleine, B. (1996). Bidirectional instrumental conditioning. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology,* 49B(4), 289-306.

Dickinson, A., Squire, S., Varga, Z., & Smith, J. W. (1998). Omission learning after instrumental pretraining. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology,* 51B(3), 271-286.

Dickinson, A., Smith, J., & Mirenowicz, J. (2000). Dissociation of pavlovian and instrumental incentive learning under dopamine antagonists. *Behavioral Neuroscience, 114*(3), 468-483.

Dickinson, A., Wood, N., & Smith, J.W. (2002). Alcohol seeking by rats: action or habit? *Quarterly Journal of Experiment Psychology: Comparative and Physiological Psychology*. 55(4), 331-348.

Dixon, A.L., Prior, M., Morris, P.M., Shah, Y.B., Joseph, M.H., & Young, A.M. (2005). Dopamine antagonist modulation of amphetamine response as detected using pharmacological MRI. *Neuropharmacology*, 48(2), 236-245.

Dunn, M.J. (2003). Effect of pharmacological manipulation of dopamine (DA) on conditional discrimination performance: implications for an impairment of conditional processing in schizophrenia. Thesis (Ph.D.) – UWC.

Dunn, M.J., Futter, D., Bonardi, C., & Killcross, S. (2005). Attenuation of d-amphetamine-induced disruption of conditional discrimination performance by α–flupenthixol. *Psychopharmacology,* 177, 296-306.

Ettinger, R.H., Reid, A.K., & Staddon, J.E.R. (1987). Sensitivity to molar feedback funcitons – a test of molar optimaility theory. *Journal of Experimental Psychology: Animal Behavior Processes,* 13(4), 366-375.

Everitt, B. J., Dickinson, A., & Robbins, T. W. (2001). The neuropsychological basis of addictive behaviour. *Brain Research Reviews, 36*(2-3), 129-138.

Everitt, B. J., & Wolf, M. E. (2002). Psychomotor stimulant addiction: a neural systems perspective. *Journal of Neuroscience,* 22(9), 3312-3320.

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience*, 8(11), 1481-1489.

Eyny, Y.S., & Horvitz, J.C. (2003). Opposing roles of D1 and D2 receptors in appetitive conditioning. *Journal of Neuroscience,* 23(5), 1584-7.

Faure, A., Haberland, U., Conde, F., & El Massioui, N. (2005). Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *Journal of Neuroscience, 25*(11), 2771-2780.

Featherstone, R. E., & McDonald, R. J. (2004). Dorsal striatum and stimulus-response learning: lesions of the dorsolateral, but not dorsomedial, striatum impair acquisition of a stimulus-response-based instrumental discrimination task, while sparing conditioned place preference. *Neuroscience,* 124, 23-31.

Fletcher, P. J. (1995). Effects of d-fenfluramine and metergoline on responding for conditioned reward and the response potentiating effect of nucleus-accumbens d-amphetamine. *Psychopharmacology, 118*(2), 155-163.

Fletcher, P. J. (1996). Injection of 5-HT into the nucleus accumbens reduces the effects of d-amphetamine on responding for conditioned reward. *Psychopharmacology, 126*(1), 62-69.

Foltin, R. W. (2004). Effects of amphetamine, dexfenfluramine, and diazepam on responding during extinction in nonhuman primates. *Pharmacology Biochemistry and Behavior, 79*(2), 325-330.

Fowler, S.C., & Liou, J.R. (1994). Microcatalepsy and disruption of forelimb usage during operant behavior: differences between dopamine D1 (SCH-23390) and D2 (raclopride) antagonists. *Psychopharmacology,* 115(1-2), 24-30.

Franken, I.H., Stam, C.J., Hendriks, V.M., & van den Brink, W. (2003). Neurophysiological evidence for abnormal cognitive processing of drug cues in heroin dependence. *Psychopharmacology*, 70(2), 205-212.

Fuchs, R.A., & Branham, R.K., & See, R.E. (2006). Different neural substrates mediate cocaine seeking after abstinence versus extinction training: a critical role for the dorsolateral caudate-putamen. *Journal of Neuroscience,* 26(13), 3584-3588.

Gerdeman, G.L., Partridge, J.G., Lupica, C.R, & Lovinger, D.M. (2003). It could be habit forming: drugs of abuse and striatal synaptic plasticity. *Trends in Neurosciences,* 26(4), 184-192.

Grant. S., Contoreggi, C., & London, E.D. (2000). Drug abusers show impaired performance in a laboratory test of decision making. *Neuropsychologia,* 38, 1180–1187.

Gray, T., & Wise, R. A. (1980). Effects of pimozide on lever pressing behavior maintained on an intermittent reinforcement schedule. *Pharmacology Biochemistry and Behavior, 12*(6), 931-935.

Graybiel, A. M. (1990). Neurotransmitters and Neuromodulators in the Basal Ganglia. *Trends in Neurosciences, 13*(7), 244-254.

Graybiel, A. M. (1995). Building action repertoires: memory and learning functions of the basal ganglia. *Current opinion in Neurobiology,* 5, 733-741.

Graybiel, A. M. (1998). The basal ganglia and chunking of actions repetoires. *Neurobiology of learning and memory,* 70, 119-136.

Graybiel, A. M., & Rauch, S. L. (2000). Toward a neurobiology of obsessive-compulsive disorder. *Neuron, 28*(2), 343-347.

Graybiel, A. M. (2004). Network-level neuroplasticity in cortico-basal ganglia pathways. *Parkinsonism and Related Disorders,* 10, 293-296.

Groenewegen, H.J., Berendse, H.W., Meredith, G.E., Haber, S.N., Voorn, S.N., Wolters, J.G., & Lohman, A.H.M. (1991). Functional anatomy of the ventral, limbic-system innervated striatum. In Willner, P & Scheel-Krüger, J. (eds). *The mesolimbic dopamine system: from motivation to action.* Pp. 1-39. Chichester: John Wiley.

Haddon, J.E., & Killcross, S. (2006). Prefrontal lesions disrupt the contextual control of response conflict. *Journal of Neuroscience,* 26(11), 2933-2940.

Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience, 20*(6), 2369-2382.

Harmer, C. J., & Phillips, G. D. (1998). Enhanced appetitive conditioning following repeated pre-treatment with d-amphetamine. *Behavioural Pharmacology, 9*(4), 299-308.

Harmer, C. J., & Phillips, G. D. (1999). Enhanced conditioned inhibition following repeated pre-treatment with d-amphetamine. *Psychopharmacology, 142*(2), 120-131.

Hay, J.F., Moscovitch, M., & Levine, B. (2002). Dissociating habit and recollection: evidence from Parkinson's disease, amnesia and focal lesion patients. *Neuropsychologia,* 40(8), 1324-34.

Hernandez, L., & Hoebel, B.G. (1988). Food reward and cocaine increase extra-cellular dopamine in the nucleus accumbens as measured by dialysis. *Life Sciences,* 42, 1705-1712.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior,* 13, 243-266.

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

Holland, P. C., (1979). Differential effects of omission contingencies on various components of Pavlovian appetitive responding in rats. *Journal of Experimental Psychology: Animal Behaviour Processes, 5,* 178-193.

Holland, P. C. (1998). Amount of training affects associatively-activated event representation. *Neuropharmacology, 37,* 461-469.

Holland, P. C. (2005). Amount of training effects in representation-mediated food aversion learning: No evidence of a role for associability changes. *Learning and Behavior,* 33(4), 464-478.

Horvitz, J.C. (2001). The effects of D1 and D2 receptor blockade on the acquisition and expression of a conditioned appetitive response. *Appetite*, 37(2), 119-120.

Howell, D. C. (2002). *Statistical methods for psychology.* Pacific Grove, CA: Duxbury.

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

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

Ito, R., Dalley, J. W., Robbins, T. W., & Everitt, B. J. (2002). Dopamine Release in the Dorsal Striatum during Cocaine-Seeking Behavior under the Control of a Drug-Associated Cue. *Journal of Neuroscience., 22*(14), 6247-6253.

James, W. (1890). *The principles of Psychology.* Macmillan and Co, Ltd.: London 1901.

Jaworski, J.N., Gonzales, R.A., & Randall, P.K. (2001). Effect of dopamine D2/D3 receptor antagonist sulpiride on amphetamine-induced changes in striatal extracellular dopamine. *European Journal of Pharmacology,* 418(3), 201-216.

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

Joel, D., & Weiner, I. (2000). The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organisation of the striatum. *Neuroscience,* 96(3), 451-474.

Jog, M. S., Kubota, Y., Connolly, C. I., Hillegaart, V., & Graybiel, A. M. (1999). Building neural representations of habits. *Science, 286*(5445), 1745-1749.

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

Jeuptner, M., Frith, C. D., Brooks, D. J., Frackowiak, R. S. J., & Passingham, R. E. (1997b). Anatomy of motor learning: II. Subcortical structures and learning by trial and error. *Journal of Neurophysiology,* 77, 1325-1337.

Karper PE, De la Rosa H, Newman ER, Krall CM, Nazarian A, McDougall SA, Crawford CA. (2002). Role of D1-like receptors in amphetamine-induced behavioral sensitization: a study using D1A receptor knockout mice. *Psychopharmacology,* 159(4), 407-414.

Kebabian, A. B., & Calne, D.B. (1979). Multiple receptors for dopamine. *Nature,* 277, 93-96.

Kennerley, S.W., Walton, M.E., Behrens, T.E.J., Buckley, M.J., & Rushworth, M.F.S. (2006). Optimal decision making and the anterior cingulated cortex. *Nature Neuroscience,* 9(7), 940-947.

Kerr, J.N., & Wickens, J.R. (2001). Dopamine D-1/D-5 receptor activation is required for long-term potentiation in the rat neostriatum in vitro. *Journal of Neurophysiology*, 85, 117-124.

Killcross, A. S., & Blundell, P. (2002). Associative representations of emotionally significant outcomes. In: Moore S & Oaksford M (Eds.) *Emotional Cognition (Advances in Consciousness Research).* John Benjamins, Amsterdam.

Killcross, S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral Cortex, 13*(4), 400-408.

Kimble, G. A., & Perlmutter, L. C. (1970). The problem of volition. *Psychological Review,* 77, 361-384.

Knowlton, B. J., Squire, L. R., Paulsen, J. R., Swerdlow, N., & Swenson, M. (1996a). Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology,* 10, 1-11.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996b). A neostriatal habit learning system in humans. *Science, 273*(5280), 1399-1402.

Koepp, M. J., Gunn, R. N., Lawrence, A.D., Cunningham, V.J., Dagher, A., Jones, T., Brooks, D.J., Bench, C.J., & Grasby, P.M. (1998). Evidence for striatal dopamine release during a video game. *Nature*, 393(6682), 266-268.

Kumari V, Corr PJ, Mulligan OF, Cotter PA, Checkley SA, Gray JA. (1997). Effects of acute administration of d-amphetamine and haloperidol on procedural learning in man. *Psychopharmacology*, 129(3), 271-276.

Kurylo, D. D., & Tanguay, S. (2003). Effects of quinpirole on behavioral extinction. *Physiology & Behavior, 80*(1), 1-7.

Lauwereyns, J., Takikiawa, Y., Kawagoe, R., Kobayashi, S., Koizumi, M., Coe, B., Sakagami, M., & Hikosaka, O. (2002). Feature-based anticipation of cues that predict reward in monkey caudate nucleus. *Neuron,* 33, 463-473.

Leckman, J. F., & Riddle, M. A. (2000). Tourette's Syndrome: when habit-forming systems form habits of their own? *Neuron, 28*(2), 349-354.

Legault, G., Smith, C.T., & Beninger, R. J. (2006). Post-training intra-striatal scopolamine or flupenthixol impairs radial maze learning in rats. *Behavioural Brain Research*, 170(1), 148-55.

Lehericy, S., Benali, H., Van de Moortele, P.F., Pelegrini-Issac, M., Waechter, T., Ugurbil, K., & Doyon, J. (2005). Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *Proceedings of the National Academy of Sciences,* 102(35), 12566-71.

Letchworth, S.R, Nader, M.A., Smith, H.R., Friedman, D.P., & Porrino, L.J. (2001). Progression of changes in dopamine transporter binding site density as a result of cocaine self-administration in rhesus monkeys. *Journal of Neuroscience,* 21(8), 2799-2807.

Ljungberg, T., Apicella, P., & Schultz, W. (1992). Responses of monkey dopamine neurons during the learning of behavioral reactions. *Journal of Neurophysiology,* 67, 145-163.

Li, Y., Kolb, B., & Robinson, T.E. (2003). The location of persistent amphetamine-induced changes in the density of dendritic spines on medium spiny neurons in the

nucleus accumbens and caudate-putamen. *Neuropsychopharmacology*, 28, 1082-1085.

Lubman, D.I., Peters, L.A., Mogg, K., Bradley, B.P., & Deakin, J.F. (2000). Attentional bias for drug cues in opiate dependence. *Psychological Medicine*, 30(1), 169-175.

Mackintosh, N. J. (1974). *The psychology of animal learning.* New York: Academic Press.

Mackintosh, N. J. (1975). A theory of attention: variations in the associability of stimulus with reinforcement. *Psychological Review,* 82, 276-298.

Mackintosh, N. J. (1983). *Conditioning and associative learning.* Oxford: Oxford University Press.

Maier, S. F., & Seligman, M.E.P. (1976). Learned helplessness: Theory and evidence. *Journal of Experimental Psychology: General,* 105, 3-46.

Mark, G.P., Smith, S.E., Rada, P.V., & Hoebel, B.G. (1994). An appetitively conditioned taste elicits a preferential increase in mesolimbic dopamine release. *Pharmacology, biochemistry, and behavior,* 48, 651-660.

McDonald, R. J., & White, N. M. (1993). A triple dissociation of memory systems: hippocampus, amygdala and dorsal striatum. *Behavioral Neuroscience,* 107, 3-22.

McDonald, R. J., & White, N. M. (1994). Parallel information processing in the water maze: evidence for independent memory systems involving the dorsal striatum and hippocampus. *Behavioral and Neural Biology,* 61, 260-270.

McDonald, R. J., King, A.L., & Hong, N.S. (2001). Context-specific interference on reversal learning of a stimulus-response habit. *Behavioural Brain Research,* 121, 149-165.

McDougall, S.A., Reichel, C.M., Cyr, M.C., Karper, P.E., Nazarian, A., & Crawford C.A. (2005). Importance of D(1) receptors for associative components of amphetamine-induced behavioral sensitization and conditioned activity: a study using D(1) receptor knockout mice. *Psychopharmacology*, 183(1), 20-30.

Mingote, S., Weber, S.M., Ishiwari, K., Correa, M., & Salamone, J.D. (2005). Ratio and time requirements on operant schedules: effort related effects of nucleus accumbens dopamine depletions. *European Journal of Neuroscience,* 21(6), 1749-1757.

Miles, F.J., Everitt, B.J., & Dickinson, A. (2003). Oral cocaine seeking by rats: action or habit? *Behavioral Neuroscience*, 117(5), 927-938.

Milner, B., Squire, L. R., & Kandel, E. (1998). Cognitive neuroscience and the study of memory. *Neuron,* 20, 445-468.

Mirenowicz, J., & Schultz, W. (1994). Importance of unpredictability for reward responses in primate dopamine neurons. *Journal of Neurophysiology,* 72, 1024-1027.

Mishkin, M., Malamut, B., & Bachevalier, J. (1984) Memories and habits: two neural systems. In Neurobiology of learning and memory (Lynch g, McGaugh J, Weinberger N., M, eds), 65-77. New York: Guildford.

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

Miyachi, S., Hikosaka, O., & Lu, X. (2002). Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Experimental Brain Research*, 146(1), 122-126.

Monterosso, J.R., Aron, A.R., Cordova, X., Xu, J., & London, E.D. (2006). Deficits in response inhibition associated with chronic methamphetamine abuse. *Drug and Alcohol Dependence*, 79(2), 273-277.

Morris, G., Arkadir, D., Nevet, A.,Vaadia, E., & Bergman, H. (2004). Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron,* 43, 133-143.

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., & Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. *Nature Neuroscience,* 9(8), 1057- 1063.

Nader, M.A., & Czoty, P.W. (2005). PET imaging of dopamine D2 receptors in monkey models of cocaine abuse: genetic predisposition versus environmental modulation. *American Journal of Psychiatry,* 162(8), 1473-1482.

Nader, M.A., Morgan, D., Gage, H.D., Nader, S.H., Calhoun, T.L., Buchheimer, N., Ehrenkaufer, R., & Mach, R.H. (2006). PET imaging of dopamine D2 receptors during chronic cocaine self-administration in monkeys. *Nature Neuroscience,* 9(8), 1050-1056.

Nakajima, S. (1986). Suppression of operant responding in the rat by dopamine D1 receptor blockade with SCH23390. *Physiological Psychology,* 14, 111-114.

Nelson, A., & Killcross, S. (2006). Amphetamine exposure enhances habit formation. *Journal of Neuroscience*, 26 (14), 3805-3812.

Niv, Y., Daw, N.D., Joel, D., & Dayan, P. (2006). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology,* (published on line 10th October, 2006).

Nordquist, R.E., Voorn, P., de Mooij-van Malsen, J.G., Joosten, R.N., Pennartz, C.M., & Vanderschuren, L.J. (2005). Augmented reinforcer value and accelerated habit formation after amphetamine sensitization. *Behavioural Pharmacology,* 16, Suppl. 1.

Ostlund, S. B., & Balleine, B.W. (2005). Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. *Journal of Neuroscience,* 25(34), 7763-70.

Packard, M. G., & White, N. M. (1991). Dissociation of Hippocampus and Caudate-Nucleus Memory-Systems by Posttraining Intracerebral Injection of Dopamine Agonists. *Behavioral Neuroscience, 105*(2), 295-306.

Packard, M. G., & McGaugh, J. L. (1994). Quinpirole and D-Amphetamine Administration Posttraining Enhances Memory on Spatial and Cued Discriminations in a Water Maze. *Psychobiology, 22*(1), 54-60.

Packard, M. G., & Teather, L. A. (1998). Amygdala modulation of multiple memory systems: hippocampus and caudate-putamen. *Neurobiology of Learning and Memory,* 69, 163-203.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience, 25*, 563-593.

Paulsen, P. J., Butters, N., Salmon, D. P., Heindel, W. C., & Swenson, M. R. (1993). Prism adaptation in Alzheimer's and Huntington's disease. *Neuropsychology,* 7, 73-81.

Pearce, J.M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review,* 87, 532-552.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., & Frith, C.D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature,* 442(7106), 1042-1045.

Phillips, A. G., & Fibiger, H. C. (1979). Decreased resistance to extinction after haloperidol: implications for the role of dopamine in reinforcement. *Pharmacology Biochemistry and Behavior, 10*(5), 751-760.

Phillips, A. G., & Fibiger, H. C. (1979). Decreased resistance to extinction after haloperidol: implications for the role of dopamine in reinforcement. *Pharmacology Biochemistry and Behavior, 10*(5), 751-760.

Phillips, G. D., Harmer, C. J., & Hitchcott, P. K. (2002). Blockade of sensitisation-induced facilitation of appetitive conditioning by post-session intra-amygdala nafadotride. *Behavioural Brain Research, 134*(1-2), 249-257.

Poldrack, R.A., Sabb, F.W., Foerde, K., Tom, S.M., Asarnow, R.F., Bookheimer, S.Y., & Knowlton, B.J. (2005). The neural correlates of motor skill automaticity. *Journal of Neuroscience*, 25, 5356-5364.

Porrino, L.J., Lyons, D., Smith, H.R., Daunais, J.B., & Nader, M.A. Cocaine self-administration produces a progressive involvement of limbic, association, and sensorimotor striatal domains. *Journal of Neuroscience,* 24(14), 3554-3562.

Quirk, G., Russo, G., Barron, J., and Lebron, K. 2000. The role of ventromedial prefrontal cortex in the recovery of extinguished fear. *Journal of Neuroscience,* 20, 6225 -6231.

Rauch, S.L., Whalen, P. J., Savage, C.R., Curran, T., & Kendrick, A. (1997). Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. *Human Brain Mapping,* 5, 124-32.

Reading, P. J., Dunnett, S. B., & Robbins, T. W. (1991). Dissociable roles of the ventral, medial and lateral striatum on the acquisition and performance of a complex visual stimulus-response habit. *Behavioural Brain Research,* 45, 147-161.

Reason, K. (1979). Actions not as planned: The price of automatisation. In G. Underwood & R. Stevens (Eds.), *Aspects of consciousness: Vol. 1. Psychological Issues.* London: Academic Press.

Reed, P. (1991). Multiple determinants of the effects of reinforcement magnitude on free-operant response rates. *Journal of the Experimental Analysis of Behavior,* 55(1), 109-123.

Rescorla, R. A., & Solomon, R.L. (1967). Two-process theory learning theory: Relationships between Pavlovian conditioning and instrumental learning. *Psychological Review,* 74, 151-182.

Rescorla, R. A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, A.H., & Prokasy, W.F. (Eds.) *Classical conditioning II: Current research and theory* (pp.64-99). New York: Appeleton-Century-Crofts.

Reynolds, J. N., Hyland, B. I., Wickens, J. R. (2001) A cellular mechanism of reward-related learning. *Nature, 413*(6851), 67-70

Rhodes, S. E. V., & Killcross, S. (2004). Lesions of rat infralimbic cortex enhance recovery and reinstatement of an appetitive Pavlovian response. *Learning and Memory,* 11(5), 611-616.

Ridley, R. M. (1994). The psychology of perserverative and stereotyped behaviour. *Progress in Neurobiology,* 44(2), 221-231.

Robbins, T. W., Giardini, V., Jones, G. H., Reading, P., & Sahakian, B. J. (1990). Effects of dopamine depletion from the caudate-putamen and nucleus-accumbens-septi on the acquisition and performance of a conditional discrimination Task. *Behavioural Brain Research, 38*(3), 243-261.

Robbins, T. W., & Everitt, B. J. (1999). Drug addiction: bad habits add up. *Nature, 398*(6728), 567-570.

Robbins, T.W. (2005). Role of cortical and striatal dopamine in cognitive fucntion. In Dunnett, S.B., Bentivoglio, M., Björklund, A., & Hökfelt, T. (Eds.) *Handbook of Chemical Neuroanatomy, Volume 21: Dopamine.* Amsterdam: Elsevier.

Robinson, T.E., & Becker, J.B. (1986). Enduring changes in brain and behaviour produced by chronic amphetamine administration: a review and analysis. *Brain Research,* 396, 157-198.

Robinson, T.E., & Berridge, K.C. (1993). The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Research Review,* 18, 247-291.

Robinson, T. E., & Kolb, B. (2004). Structural plasticity associated with exposure to drugs of abuse. *Neuropharmacology, 47,* 33-46.

Rogers, R.D., Everitt, B.J., Baldacchino, A., Blackshaw, A.J., Swainson, R., Wynne, K., Baker, N.B., Hunter, J., Carthy, T., Booker, E., London, M., Deakin, J.F.W., Sahakian, B.J., & Robbins, T.W. (1999). Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: Evidence for monoaminergic mechanisms. *Neuropsychopharmacology,* 20(4), 322-339.

Roncacci, S., Troisi, E., Carlesimo, G. A., Nocentini, U., & Caltagirone, C. (1996). Implicit memory in Parkinsonian patients: evidence for deficient skill learning. *European Journal of Neurology,* 36, 154–159.

Saint-Cyr, J.A., Taylor, A.E., & Lang, A.E. (1988) Procedural learning and neostriatal dysfunction in man. *Brain*, 111, 941-959.

Saka, E., Goodrich, C., Harlan, P., Madras, B., & Graybiel, A.M. (2004). Repetitive behaviours in monkeys are linked to specific striatal activation patterns. *Journal of Neuroscience,* 24(34), 7557-7565.

Salamone, J.D., Steinpreis, R.E., McGullough, L.D., Smith, P., Grebel, D., & Mahan, K. (1991). Haloperidol and nucleus accumbens dopamine depletion suppress lever pressing for food but increase free food consumption in a novel food choice procedure. *Psychopharmacology,* 104(4), 515-521.

Salamone, J.D., Cousins, M.S., Bucher, S. (1994). Anhedonia or anergia? Effects of haloperidol and nucleus accumbens depletion on instrumental response selection in a T-maze cost/benefit procedure. *Behavioural Brain Research,* 65, 221-229.

Salamone, J.D., Wisniecki, A., Carlson, B.B., & Correa, M. (2001). Nucleus accumbens dopamine depletions make animals highly sensitive to high fixed ratio requirements but do not impair primary food reinforcement. *Neuroscience,* 5(4), 863-870.

Salmon, D. P., & Butters, N. (1995). Neurobiology of skill and habit learning. *Current Opinion in Neurobiology,* 5, 184-190.

Salo, R., Nordahl, T.E., Possin, K., Leamon, M., Gibson, D.R., Galloway, G.P., Flynn, N.M., Henik, A., Pfefferbaum, A., & Sullivan, E.V. (2002). Preliminary evidence of reduced cognitive inhibition in methamphetamine-dependent individuals. *Psychiatry Research*, 111(1), 65-74.

Schoenbaum, G., & Setlow, B. (2005). Cocaine makes actions insensitive to outcomes but not extinction: Implications for altered orbitofrontal-amygdalar function. *Cerebral Cortex, 15*(8), 1162-1169.

Schultz, W. (1998). Predicitve reward signal of dopamine neurons. *Journal of Neurophysiology,* 80, 1-27.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron,* 36, 241-263.

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *Journal of Neuroscience,*. 13, 900-913.

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

Schultz, W., Tremblay, L., & Hollerman, J.R. (2003). Changes in behaviour-related neuronal activity in the striatum during learning. *Trends in Neurosciences,* 26(6), 321-328.

Sharf, R., Lee, D.Y., & Ranaldi, R. (2005). Microinjections of SCH 23390 in the ventral tegmental area reduce operant responding under a progressive ratio schedule of food reinforcement in rats. *Brain Research,* 1033(2), 179-85.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review,* 84, 127-190.

Simon, S.L., Domier, C., Carnell, J., Brethen, P., Rawson, R., & Ling, W. (2000). Cognitive impairment in individuals currently using methamphetamine. *The American Journal on Addictions*, 9(3):222-231.

Smith, J.K., Neill, J.C., Costall, B. (1997). Bidirectional effects of dopamine D2 receptor antagonists on responding for a conditioned reinforcer. *Pharmacology, Biochemistry and Behavior*, 57(4), 843-849.

Sofuoglu, M., & Kosten, T.R. (2005). Novel approaches to the treatment of cocaine addiction. *CNS Drugs,* 19(1), 13-25.

Spence, K. W. (1956). *Behavior theory and conditioning.* New Haven, CT: Yale University Press.

Stewart, J., &. Vezina, P. (1989). Microinjections of SCH-23390 into the ventral tegmental area and substantia nigra pars reticulata attenuate the development of sensitization to the locomotor activating effects of systemic amphetamine. *Brain Research*, 495(2), 401-406.

Stewart, J., & Badiani, A. (1993). Tolerance and sensitization to the behavioral effects of drugs. *Behavioural Pharmacology, 4,* 289-312.

Sutton, R. S., & Barto, A .G. (1998). *Reinforcement learning: an introduction.* Cambridge, MA: MIT Press.

Takagishi, M., & Chiba, T. (1991). Efferent projections of the infralimbic cortex (area 25) region of the medial prefronatal cortex in the rar: an anterograde traces PHA-L study. *Brain Research,* 566, 26-39.

Tanabe, L. M., Suto, N., Creekmore, E., Steinmiller, C.L., & Vezina, P. (2004). Blockade of D2 dopamine receptors in the VTA induces a long-lasting enhancement of the locomotor activating effects of amphetamine. *Behavioural Pharmacology,* 15(5-6), 387-395.

Taylor, J. R., & Horger, B. A. (1999). Enhanced responding for conditioned reward produced by intra- accumbens amphetamine is potentiated after cocaine sensitization. *Psychopharmacology, 142*(1), 31-40.

Taylor, J. R., & Jentsch, J. D. (2001). Repeated intermittent administration of psychomotor stimulant drugs alters the acquisition of pavlovian approach behavior in rats: Differential effects of cocaine, d-amphetamine and 3,4-methylenedioxymethamphetamine ("ecstasy"). *Biological Psychiatry, 50*(2), 137-143.

Tombaugh, T.N., Tombaugh, J., & Anisman, H. (1979). Effects of dopamine receptor blockade on alimentary behaviors: home cage food consumption, magazine training, operant acquisition, and performance. *Psychopharmacology,* 66(3), 219-225.

Thorndike, E.L., (1911). *Animal intelligence: Experimental studies.* New York: Macmillan.

Tiffany, S.T. (1990). A cognitive model of drug urges and drug-use behavior: Role of automatic and nonautomatic processes. *Psychological Review,* 97(2), 147-168.

Tolman, E. C., (1932). *Purposive behaviour in animals and man.* New York: Century.

Trapold, M. A., & Overmier, J. B. (1972). The second learning process in instrumental learning. In A. A. Black & W. F. Prakasy (Eds.), *Classical conditioning: II. Current research and theory.* New York: Appleton-Century-Crofts.

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

Van den Bos, R., van der Harst, J., Jonkman, S., Schilders, M., & Spruijt, B. (2006). Rats assesses costs and benefits according to an internal standard. *Behavioural Brain Research,*

Vanderschuren, L., & Kalivas, P. W. (2000). Alterations in dopaminergic and glutamatergic transmission in the induction and expression of behavioral sensitization: a critical review of preclinical studies. *Psychopharmacology, 151*(2-3), 99-120.

Varvel, S.A., Vann, R.E., Wise, L.E., Philibin, S.D., & Porter, J.H. (2002). Effects of antipsychotic drugs on operant responding after acute and repeated administration. *Psychopharmacology, 160*(2), 182-191.

Vezina, P. (1996). D1 dopamine receptor activation is necessary for the induction of sensitization by amphetamine in the ventral tegmental area. *Journal of Neuroscience,* 16(7), 2411-2420.

Vezina, P., & Stewart, J. (1989). The effect of dopamine receptor blockade on the development of sensitization to the locomotor activating effects of amphetamine and morphine. *Brain Research,* 499(1), 108-120.

Volkow, N.D., Wang, G.J., Fowler, J.S., Logan, J., Gatley, S.J., Gifford, A., Hitzemann, R., Ding, Y.S., & Pappas, N. (1999). Prediction of reinforcing responses to psychostimulants in humans by brain dopamine D2 receptor levels. *The American Journal of Psychiatry*, 156(9), 1440-1443.

Volkow, N.D., Chang, L., Wang, G.J., Fowler, J.S., Ding, Y.S., Sedler, M., Logan, J., Franceschi, D., Gatley, J., Hitzemann, R., Gifford, A., Wong, C., & Pappas, N. (2001). Low level of brain dopamine D2 receptors in methamphetamine abusers: association with metabolism in the orbitofrontal cortex. *American Journal of Psychiatry,* 158(12), 2015-2021.

Volkow, N.D., Fowler, J.S., Wang, G.J., & Swanson, J.M. (2004). Dopamine in drug abuse and addiction: results from imaging studies and treatment implications. *Molecular Psychiatry,* 9(6), 557-569.

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature,* 412(6842), 43-48.

Walton, M.E., Bannerman, D.M., & Rushworth, M.F.S. (2002). The role of the rat medial frontal cortex in effort-based decision making. *Journal of Neuroscience,* 22(24), 10996-11003.

Wamsley, J.K., Gehlert, D.R., Filloux, F.M., & Dawson, T.M. (1989). Comparison of the distribution of D-1 and D-2 dopamine receptors in the rat brain. *Journal of Chemical Neuroanatomy*, 2(3), 119-37.

Wang, G.J, Volkow, N.D., Fowler, J.S., Logan, J., Abumrad, N.N., Hitzemann, R.J., Pappas, N.S., & Pascani, K. (1997). Dopamine D2 receptor availability in opiate-dependent subjects before and after naloxone-precipitated withdrawal. *Neuropsychopharmacology,* 16(2), 174-182.

Wang, G.J., Volkow, N.D., Fowler, J.S., Cervany, P., Hitzemann, R.J., Pappas, N.R., Wong, C.T., & Felder, C. (1999). Regional brain metabolic activation during craving elicited by recall of previous drug experiences. *Life Sciences,* 64(9), 775-784.

Watanabe, K., & Kimura, M. (1998). Dopamine receptor-mediated mechanisms involved in the expression of learned activity of primate striatal neurons. *Journal of Neurophysiology,* 79, 2568-2580.

White, N. M., & McDonald, R. J. (2002). Multiple parallel memory systems in the brain of the rat. *Neurobiology of Learning and Memory*, 77, 125-184.

Wickens, J.R., Reynolds, J.N., & Hyland, B.I. (2003). Neural mechanisms of reward-related motor learning. *Current Opinion in Neurobiology*, 13(6), 685-690.

Wise, R. A. (1998). Drug activation of brain reward pathways. *Drug and Alcohol Dependence,* 51, 12-22.

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews Neuroscience, 5*(6), 483-494.

Wise, R.A., & Schwartz, H.V. (1981). Pimozide attenuates acquisition of lever-pressing for food in rats. *Pharmacology, biochemistry, and behaviour*, 15(4), 655-656.

Wise, R.A., & Bozarth, M. A. (1987). A psychomotor stimulant theory of addiction. *Psychological Review*, 94(4), 469-492.

Wyvell, C. L., & Berridge, K. C. (2000). Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: enhancement of reward 'wanting' without enhanced 'liking' or response reinforcement. *Journal of Neuroscience, 20*(21), 8122-8130.

Wyvell, C. L., & Berridge, K. C. (2001). Incentive sensitization by previous amphetamine exposure: increased cue-trigged 'wanting' for sucrose reward. *Journal of Neuroscience, 21*(19), 7831-7840.

Xu, M., Moratalla, R., Gold, L.H., Hiroi, N., Koog, G.F., Graybiel, A.M., & Tonegawa, S. (1994). Dopamine D1 receptor mutant mice are deficient in striatal expression of dynorphin and in dopamine-mediated behavioural responses. *Cell,* 79(4), 729-742.

Xu, M., Guo, Y.H., Vorhees, C.V., & Zhang, J.H. (2000). Behavioural reponses to cocaine and amphetamine administration in mice lacking the dopamine D1 receptor. *Brain Research,* 852(1), 198-207.

Yamamoto. Y., Nakanishi, H., Takai, N., Shimazoe, T., Watanabe, S., & Kita, H. (1999). Expression of N-methyl-D-aspartate receptor-dependent long-term potentiation in the neostriatal neurons in an in vitro slice after ethanol withdrawal of the rat. *Neuroscience,* 91, 59-68.

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

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience, 19*(1), 181-189.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005a). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience, 22*(2), 513-523.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2005b). Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *European Journal of Neuroscience, 22*(2), 505-512.

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

Yin, H.H., Zhuang, X., & Balleine, B.W. (2006b). Instrumental learning in hyperdopaminergic mice. *Neurobiology of Learning and Memory,* 85(3), 283-288.

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience,* 7, 464-476.

Yue, E.L., Cleland, T.A., Pavlis, M., & Linster, C. (2004). Opposing effects of D-1 and D-2 receptor activation on odor discrimination learning. *Behavioral Neuroscience,* 118(1), 184-190.

Zhang, D.S., Zhang, L., Tang, Y., Zhang, Q., Lou, D.W., Sharp, F.R., Zhang, J.H., & Xu, M. (2005). Repeated cocaine administration induces gene expression changes through the dopamine D1 receptors. *Neuropsychopharmacology,* 30(8), 1443-1454.

Zink, C. F., Pagnoni, G., Martin-Skurski, M.E., Chappelow, J.C., & Berns, G.W. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron,* 42, 509-517.