PERGAMON

Neuropharmacology 37 (1998) 407–419

NEURO
PHARMACOLOGY

Goal-directed instrumental action: contingency and incentive learning and their cortical substrates

Bernard W. Balleine [a], Anthony Dickinson [b,*]

[a] *Department of Psychology, UCLA, Franz Hall, Los Angeles, CA 90095-1563, USA*

[b] *Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK*

Accepted 10 March 1998

Abstract

Instrumental behaviour is controlled by two systems: a stimulus–response habit mechanism and a goal-directed process that involves two forms of learning. The first is learning about the instrumental contingency between the response and reward, whereas the second consists of the acquisition of incentive value by the reward. Evidence for contingency learning comes from studies of reward devaluation and from demonstrations that instrumental performance is sensitive not only the probability of contiguous reward but also to the probability of unpaired rewards. The process of incentive learning is evident in the acquisition of control over performance by primary motivational states. Preliminary lesion studies of the rat suggest that the prelimibic area of prefrontal cortex plays a role in the contingency learning, whereas the incentive learning for food rewards involves the insular cortex. © 1998 Elsevier Science Ltd. All rights reserved.

Keywords: Cortex; Instrumental conditioning; Reinforcement; Reward; Rats

1. Introduction

Prediction and control are the keys to successful adaptation to varying environments. Predictive learning allows an animal to anticipate biologically important events and resources by detecting and learning about signals of their occurrence. Traditionally, this form of learning has been studied using Pavlovian or classical conditioning procedures in which the signal acquires the capacity to elicit anticipatory responses as a result of its predictive association with a reinforcer. Although the functional importance of Pavlovian responses is well established (Hollis et al., 1997), their adaptive form is determined by evolutionary processes rather than individual learning, with the consequence that a purely Pavlovian animal is at the mercy of the stability of the causal consequences of its behaviour.

This point can be illustrated by one of the simplest behavioural capacities-the ability to approach signals of valuable resources. Having fed chicks at a distinctive food bowl, Hershberger (1986) found that when they were later removed from the vicinity of the bowl, not surprisingly they immediately ran back to it. Presumably, the initial feeding established the visual features of the bowl as a Pavlovian signal for food, capable of eliciting a conditioned approach response. For a second group, however, Hershberger reversed the normal relationship between locomotion and spatial translation by placing the chicks in a 'looking glass' world where the bowl receded twice as fast as they ran towards it, and approached them at twice the speed that they ran away from it. The reversal of the normal relation between locomotion and relative spatial translation required the chicks to learn to run away from the bowl in order to reach it. This they were unable to do over 100 min of training.

The problem for Hershberger's chicks resides with their insensitivity to the change in the causal consequences of their behaviour, at least with respect to spatial locomotion, and it is the ability to learn about such causal relationships that represents a second form of acquired behavioural adaptation to varying environments. Learning about behaviour-outcome associations is typically studied using instrumental conditioning procedures in which a relationship is arranged between an

* Corresponding author: Tel.: +44 1954 333577; fax: +44 1954 333564; e-mail: ad15@cus.cam.ac.uk.

PII: S0028-3908(98)00033-1

action and a reinforcer. Whereas Pavlovian conditioning enables an animal to anticipate motivationally significant events, it is instrumental conditioning that allows control over these events in the service of its needs and desires. And it is instrumental learning that is the focus of this paper.

The classic learning theories of the neobehaviourist era (Tolman, 1932; Hull, 1943) were developed in response to studies of what was, at least nominally, instrumental conditioning. In the 1970s, however, the primary focus of learning theory shifted from the instrumental to the Pavlovian paradigm primarily for technical reasons. In the Pavlovian paradigm, the experimenter has control over the critical elements of the relationship, the signal and the reinforcer, whereas one of the elements of the instrumental association, the response itself, is under the subject's control. Consequently, contemporary theories of conditioning (Rescorla and Wagner, 1972; Wagner, 1981; Pearce and Hall, 1980; Gallistel, 1990) have focused almost exclusively on the Pavlovian paradigm. Much the same is true of the neurobiological analysis of conditioning. The favoured procedures for investigating the neural structures mediating conditioning are Pavlovian (Lavond et al., 1993; Ledoux, 1995) with the consequence that most neural network models (Hawkins and Kandel, 1984; Gluck and Thompson, 1987; Schmajuk, 1997) also address this form of learning.

It is true that brain mechanisms mediating rewards in general have been the subject of intensive study (see Robbins and Everitt, 1996, for a recent review), but this analysis has been largely undertaken without any attempt to specify exactly how reward processes make contact with structures that mediate instrumental action. To the extent that an associative structure and learning process for instrumental action has been specified, it typically takes the form of a variant of the classic stimulus–response (S–R)/reinforcement system originally advanced in Thorndike (1911) 'Law of Effect' (Donahoe et al., 1993). The central idea embodied in this so-called law is simple-the presentation of a reward shortly after the performance of an instrumental action strengthens or reinforces an association between the stimuli present when the response was performed and the response production mechanism so that these stimuli become capable of eliciting the response.

There is no doubt that the S–R/reinforcement mechanism, when embodied within artificial creatures and elaborated with attentional and motivational mechanisms, can support sophisticated and complex instrumental behaviour (Grand et al., 1996). As a system for adapting to the causal structure of the environment, however, an S–R process has two major limitations. The first relates to the fact that S–R processes are sensitive to only the contiguous pairing of action and reinforcer rather than to the causal relationship between these events. As a result, it is prone to develop superstitious responding under conditions in which a reinforcer reliably follows a response even if there is no causal association between the response and reward.

The second limitation arises from the failure of an S–R process to encode the consequences of a response or, in other words, to represent the causal relationship between an action and a reward. All that is acquired during instrumental learning, according to this theory, is a procedural connection between a stimulus and a response so that the former becomes capable of reliably eliciting the latter. As a result, an S–R agent does not 'know' about the consequences of its behaviour and thus cannot evaluate different courses of action in terms of the relevance of the outcomes to its current needs and motivational states.

In the next two sections, we describe evidence demonstrating that real animals are subject to neither of these constraints.

2. Instrumental contingency

Hammond (1980) was the first to demonstrate that animals are sensitive to the causal relation between response and reward even when the contiguous pairings between them are kept constant. Under his schedule the first response in each 1-s period has a fixed probability of being reinforced. Thus, for example, hungry rats might be trained to press a lever under a schedule in which the first press in each second is followed the delivery of a food pellet with a fixed probability. The causal relationship between lever pressing and food delivery can then be degraded by increasing the probability that a food pellet will be delivered at the end of any second in which the animal does not press the lever until, when these two probabilities are equal, pressing has no effect on the likelihood of the reward. The important feature of this way of manipulating the causal relationship between response and reinforcer is that the contingency is degraded without altering the probability that a response is paired the reward. Thus, other things being equal, S–R theory predicts that degrading the contingency in this manner should have no impact on the strength of the response.

At variance with this prediction, Hammond (1980) found that enhancing the probability of a reward in the absence of the response depresses instrumental performance. We cannot be certain, however, that this decline reflects a sensitivity to the instrumental contingency per se, because presenting unpaired rewards may well strengthen competing responses, such as approaching the food source, that interfere with lever pressing. To control for the effect of response competition, a modified procedure has been developed that employs two different rewards (Colwill and Rescorla, 1986;

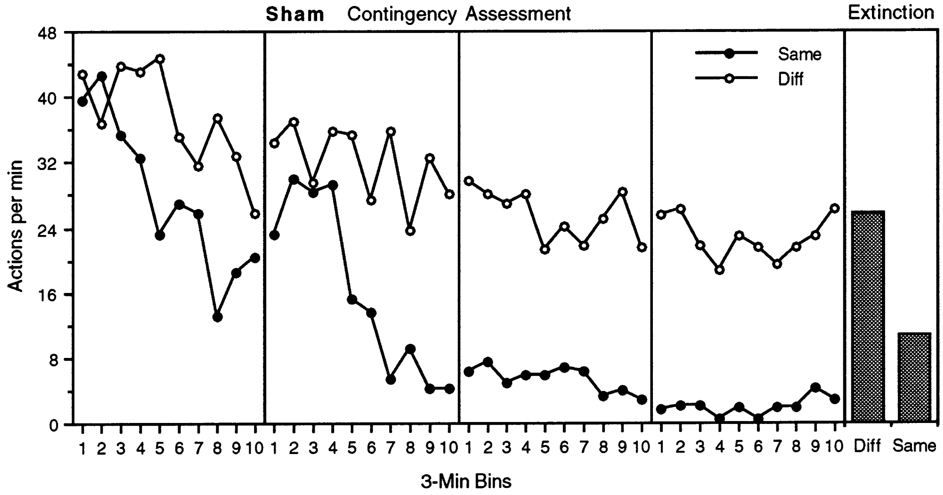

Fig. 1. The mean number of actions (lever pressing and chain pulling) per min during the four sessions of training under the non-contingent training and during the final extinction test (right-hand panel). The response rates are shown separately for the actions paired with the reward that was same as and different from the unpaired reward.

Dickinson and Mulatero, 1989; Williams, 1989). For example, we trained hungry rats to press a lever and pull a chain in separate sessions with one response reinforced by food pellets and the other by a starch solution[1]. At the end of pretraining the appropriate reward was delivered with a probability of 0.05 following each second containing at least one response. The instrumental contingency was then degraded by arranging for one of the foods, the pellets for half of the animals and the starch solution for the remainder, to be delivered following each second without a response with the same, 0.05 probability. Thus, under this non-contingent schedule the probability of a reward was the same following each second with at least one response as that following each second without a response.

This contingency manipulation has a differential impact on the causal status of the two actions. When the unpaired food is the same as the one paired with the response, this action is rendered causally ineffective-performing the action has no effect on the probability of either reward. By contrast, the action is still causally effective when the paired reward differs from the unpaired one-by responding, the animal can increase the frequency of the paired reward relative to the unpaired one. The important point about this procedure is that any difference in the performance of the two actions cannot be explained in terms of response competition. As the same unpaired reward is presented with the same probability during the performance of both actions, any response competition should be equivalent.

Fig. 1 shows that animals are sensitive to the difference in the causal status of the two actions under the non-contingent schedule. Displayed are the rates of responding across the four sessions of non-contingent training as a function of whether the unpaired reward was the same as or different from the paired reward. In the first session, both actions tended to decline but with further training the action paired with the different reward stabilized at a level above that paired with the same reward, which dropped to a very low level. The final, right-hand panel illustrates that these differences persisted into a final extinction test in which neither the paired nor unpaired rewards were presented. By the end of non-contingent training and during the extinction test all eight animals trained on this procedure performed the action paired with the reward that differed from unpaired food more than the response for which the paired and unpaired rewards were the same.

Although some of the decline observed under the non-contingent schedule may have been due to response competition, the difference between the performance of the two actions reflected a direct effect of the differing instrumental contingencies. Thus, instrumental performance is sensitive not just to the contiguity between response and reinforcer but also to the contingency between them. As S–R/reinforcement theory identifies contiguity as the critical variable, this finding lies outside the scope of the theory.

3. Reward devaluation

The second limitation of an S–R process lies with the absence of any encoding of the relationship between the response and the reward with the consequence that an S–R agent is incapable of truly goal-directed be-

[1] The strain and sex of the rats, the training apparatus and manipulanda, the rewards, the deprivation regime and the housing conditioning in this and all the other experiments reported were the same as those used in Balleine (1992).

haviour. In other words, the theory treats all instrumental behaviour as simple, elicited habits. This is certainly not the interpretation that our folk psychology gives for many of our instrumental actions. We typically regard our actions as purposive and explicitly selected and performed because of our knowledge of their beneficial consequences, and there is now good experimental evidence that we share this capacity for goal-directed, instrumental behaviour with other animals.

This point is illustrated by considering the impact of devaluing a reward on instrumental performance. For example, in one study we trained hungry rats to press a lever and pull a chain, again with one response reinforced by food pellets and the other by a starch solution in a counterbalanced assignment. As in the training phase described for the previous study, the appropriate reward was delivered with a probability of 0.05 in each second containing at least one response. No unpaired rewards were scheduled in this study. Following this training, we devalued one of the rewards by using a specific satiety procedure. It is well known that prefeeding a particular food reduces that the subsequent hedonic reactions to and consumption of that specific food relative to other, non-prefed foods (see Hetherington and Rolls, 1996, for a recent review), and so following the instrumental training we prefed the animals one of the two rewards for 1 h in their home cage. For half of the animals the prefed food was the pellets and for the remainder the starch solution. Immediately following this prefeeding, the animals were returned to the training context which now contained both the lever and the chain, and the number of lever presses and chain pulls performed were recorded. It is important to note that this test was conducted in extinction and in the absence of any rewards. As a consequence, any effect of the prefeeding could not have been mediated by altering the direct impact of the rewards themselves but must reflect knowledge of the two rewards acquired during the initial instrumental training.

Simple S–R theory predicts that any effect of the prefeeding should be the same for the two responses. This is because, according to the theory, rewarding events act merely to reinforce the connection between the S and R and are not themselves encoded. Consequently, there is no obvious route by which devaluation of one of the rewards could affect performance of one response rather than the other as long as the foods themselves are not presented during the test. As Fig. 2 illustrates, however, devaluing one of the rewards by prefeeding selectively and profoundly reduced performance of its associated response relative to the action trained with the non-prefed food that should have retained its value (see also Colwill and Rescorla, 1985a). During this extinction test, all eight rats performed the action trained with the devalued reward less than that trained with the valued reward.

S–R theorists have made valiant efforts to explain reinforcer devaluation effects without appealing to knowledge of the response-reward association. For example, it has been argued that during training the manipulanda, in this case the lever and the chain, come to elicit different Pavlovian responses because of their differential association with the two rewards. In turn, these Pavlovian responses are assumed to produce different internal stimuli to which the appropriate instrumental response for the manipulandum is conditioned by the S–R/reinforcement mechanism. Given these assumptions, an S–R account of reward devaluation can be formulated by supposing that devaluing one reward reduces the corresponding Pavlovian response, thereby weakening any internal stimulus controlling the instrumental response (Donahoe et al., 1997). This analysis ignores, however, the fact that reward devaluation effects can be observed when animals are trained to perform different actions on a single manipulandum. For example, in a study very similar to the one described above, Dickinson et al. (1996) trained rats to press a vertically suspended pole in one direction for food pellets and in the other for the starch solution. The elaborated S–R account, applied to this task, must predict that any Pavlovian responses elicited by the pole should exert comparable control over both instrumental responses and, therefore, that devaluation will affect both actions equally. In contrast to this prediction, when Dickinson et al. (1996) devalued either the

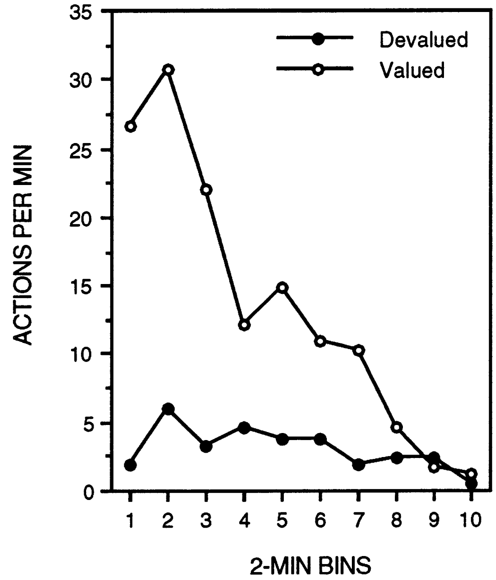

Fig. 2. The mean number of actions (lever pressing and chain pulling) per min during the extinction test following prefeeding with one of the rewards. The response rates are shown separately for the action trained with the reward that was devalued by prefeeding and for the action trained with the reward that was not prefed and therefore should have retained its value.

pellets or the starch by prefeeding, performance of the action that, in training, had delivered the subsequently prefed food was selectively reduced (Colwill and Rescorla, 1986).

In conclusion, there is compelling evidence from this and other devaluation studies (Adams and Dickinson, 1981, Colwill and Rescorla, 1985a) that instrumental action can be controlled by knowledge of the response-reward contingency in a way that lies outside the scope of even elaborated S–R theories.

4. Motivational control

The fact that a hungry rat presses a lever more rapidly for a food reward than a sated animal hardly warrants experimental demonstration, and yet the explanation of such simple motivational effects has always been problematic for theories of instrumental action. The deceptively simple explanation that the motivational state of hunger activates specifically food-directed behaviour is precluded within S–R/reinforcement theory. In the absence of any knowledge of the consequence of behaviour, an agent cannot select a response on the basis of the relevance of its reinforcer to the current motivational state. It is for this reason that classic S–R theory embraced a general drive theory of motivation (Hull, 1943) which assumes that motivational states, whether induced by food deprivation, sexual arousal, thermoregulative imbalance or whatever, act through a general state that potentiates any predominate S–R habit. Whatever selectivity is exerted by motivational states, within the theory it acts by virtue of the fact that such states have stimulus properties that can elicit responses through the standard S–R process (Davidson, 1993).

It is not always the case, however, that motivational states have a direct effect on instrumental performance. We and our colleagues (Dickinson et al., 1995) trained rats to lever press for a total of 120 food pellets that were novel for the animals and discriminably different from their maintenance diet. We ensured that the rats were hungry during training by depriving them of this maintenance diet. The left-hand panel of Fig. 3 illustrates the results of a subsequent extinction test in which the motivational state of the animals was varied. One group of animals, Group HUN, remained on the deprivation schedule and so were hungry at the time of testing, whereas the other group, Group SAT, received free access to their maintenance diet overnight and thus were in this sense sated at the time of testing. Surprisingly, the motivational state at the time of testing had no impact on instrumental performance in that animals in Group SAT pressed just as frequently as those in Group HUN.

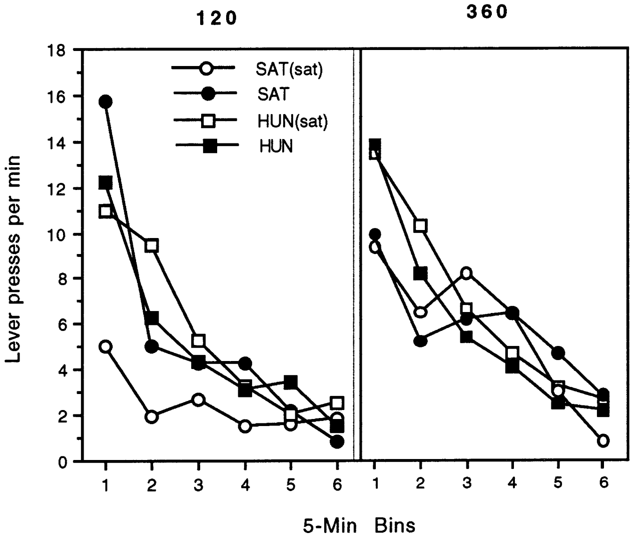

Fig. 3. Mean number of lever presses per min during the extinction test. Groups HUN and HUN(sat) were deprived of their maintenance diet and therefore hungry, whereas Groups SAT and SAT(hun) were sated on their maintenance diet. The animals in Groups SAT(sat) and HUN(sat) had received prior experience with the food reward in the sated state, whereas those in Groups SAT and HUN had only experienced the food reward previously while hungry. The left-hand panel illustrates that the performance of animal that received only 120 rewards during instrumental training and the right-hand panel the performance of animals that received 360 rewards.

In order for the motivational state at the time of testing to control performance, the animals had to receive prior experience with the food reward in the sated or non-deprived state. Between the end of instrumental training and the extinction test, two further groups were allowed to eat the food pellets in the non-deprived or sated state in separate feeding cages before being tested either hungry in the case of Group HUN(sat) or sated in the case of Group SAT(sat). These groups exhibited a very different pattern of responding. Whereas Group HUN(sat) responded at a comparable level to Groups HUN and SAT, the animals that had received prior experience with the food reward in the non-deprived state, Group SAT(sat), pressed a low rate throughout the test. Thus, it would appear that our rats had to learn that the reward pellets, which they had never previously eaten when undeprived, were less attractive when sated than when hungry.

These results accord with a number of other recent demonstrations that shifts in motivational state do not necessarily have a direct impact on instrumental performance in the manner envisaged by standard drive theory (Dickinson and Balleine, 1994, 1995). Rather prior experience with the reward in the shifted state is required if the current motivational state is to control performance, and we have interpreted such findings as evidence that animals have to learn about the significance of rewards in different motivational states through experience with them in the respective states.

5. Multiple learning processes

This brief and selective survey demonstrates that the processes mediating simple instrumental or operant behaviour are much more complex than is envisaged by classic S–R/reinforcement theory. Specifically, we have identified two features of instrumental performance that lie outside the scope of this theory. First, performance is sensitive not only to the contiguity between the response and the reward but also their contingency. Secondly, the impact of reward devaluation demonstrates that some representation of the reward is encoded in the associative structures controlling performance and, moreover, that this control depends upon the contingent relationship between the response and the reward. Taken together, these two findings suggest that, at least under certain training conditions, instrumental acquisition involves the encoding of the relationship between response and reward, a form of learning that we shall refer to as contingency learning.

In addition, studies of the impact of motivational shifts implicates a role for another learning process, incentive learning (Dickinson and Balleine, 1994, 1995). The effect of motivational shifts suggests that motivational states also act via a representation of the reward, one that encodes the incentive value of the reward and the way in which that value is modulated by differing motivational states. This representation of incentive value is acquired through consummatory experience with the reward in the respective motivational state.

This analysis should not be taken, however, to imply that an S–R/reinforcement mechanism plays no role in instrumental conditioning-it clearly does. At one time or an other we all become aware that our daily behaviour is riddled with simple habits, which are elicited and executed without thought for their consequences. Folk psychology has it that much of our behaviour starts out as intentional, goal-directed actions but that with repetitive practice these actions are transformed into S–R habits. And, indeed, this transition can be demonstrated in the conditioning laboratory. In the previous section we described a study in which a food reward was devalued by a combination of a shift from a hungry to a sated state in combination with an incentive learning treatment (see left panel of Fig. 3). An important feature of this study is that the animals received relatively limited instrumental training in that only 120 lever presses were reinforced with the food. By contrast, a second group of rats received three times as much instrumental training by being allowed to earn 360 pellets by lever pressing. The right-hand panel of Fig. 3 illustrates the performance of these animals in the extinction test. In contrast to the animals that received only 120 rewards during training, the shift to the non-deprived or sated state had a direct, albeit small effect on performance in that the rats tested in the non-deprived or sated state, Groups SAT and SAT(-sat), pressed less than those tested hungry, Group HUN and HUN(sat).

Importantly, however, the incentive learning treatment had no detectable effect after overtraining. In the 360-reward condition, Group SAT(sat), which had the prior opportunity to learn about the low value of the food pellets in the non-deprived state, pressed just as much as Group SAT, which did not have this incentive learning experience (see right-hand panel of Fig. 3). This pattern contrasts with that observed in the 120-condition in which Group SAT(sat) showed very little responding throughout the test relative to Group SAT (see left-hand panel of Fig. 3). Thus, it would appear that instrumental performance is impervious to reward devaluation after overtraining, a pattern that is indicative of control by an S–R process (Adams, 1982; Colwill and Rescorla, 1985b, 1988).

In summary, the control of simple instrumental action, such as lever pressing for food by a hungry rat, is complex and mediated by at least three different learning processes. The first is a contingency learning processes that enables the animal to encode the relation between an action and a reward. Whether or not this encoding actually represents the causal nature of this relation is probably impossible to determine in animals but the fact that instrumental learning by rats and acquisition of causal judgments of the effectiveness of actions by humans shows high concordance across many variables (Dickinson and Shanks, 1995) is certainly suggestive. The second is an incentive learning process which allows the animals to assign an appropriate value to a reward and learn how this value is modulated by its motivational states. This learning process is engaged when animals contact and experience the reward in the relevant state. Finally, there is evidence that instrumental responding can be controlled by the classic S–R mechanism which appears to predominate after more extended training.

6. Cortical structures and instrumental action

A favoured strategy in the psychological study of learning and memory is that of process dissociation, and the present tripartite analysis is no exception. Indeed, our analysis of instrumental learning conforms to the popular distinction between declarative and procedural learning (Dickinson, 1980; Squire, 1992) with contingency learning being declarative in nature and habit learning procedural. Such distinctions are first and foremost psychological in that they stand or fall by purely behavioural data. The hope is, however, that the neurobiology of behaviour is transparent with respect to such psychological distinctions in the sense that the different forms of learning can be mapped onto the underlying brain systems. In this hope, we have used our behavioral analysis to guide out research into the brain systems mediating instrumental action.

We suspect that the capacity for S–R learning is widely distributed throughout the central nervous system. Thus, for example, Wolpaw and his colleagues (Wolpaw et al., 1989; Chen and Wolpaw, 1995) have shown that the spinal reflexes can be modified by instrumental contingencies and, furthermore, provided evidence that the site of plasticity lies within the chord itself (Carp and Wolpaw, 1994). Further, on the basis of its intrinsic structure and connectivity with other brain areas, a number of authors (White, 1989) have favoured the basal ganglia as an important locus of S–R learning. And, of course, convincing arguments could be presented for the involvement of many other structures in S–R learning. Nevertheless, however distributed this learning may turn out to be, it is important to note that our tripartite model of instrumental action predicts that the expression of S–R learning will, at some level, be modulated by neural structures that support contingency and incentive learning. Although this is an important prediction of the model, its evaluation must surely await an understanding of the brain systems that mediate goal-directed action and, as a consequence, our research is focused on this issue. Given the rich projections from cortex to brain stem and basal ganglia and the modulatory role assigned to these structures (McGaugh et al., 1995), we investigated the role of cortical structures in the learning processes that mediate goal-directed action.

Without a functioning dorsolateral prefrontal cortex, humans appear to be largely 'stimulus bound' and have little confidence in their ability to interact with the environment (Knight et al., 1995). Indeed, PET studies of 'voluntary movement' in humans (Frith et al., 1991) and the delayed response task in primates (Goldman-Rakic, 1994) implicate the dorsolateral prefrontal cortex in functions such as the control of purposive action, response selection and 'planning' that appear to require the deployment of action-outcome knowledge. Although little research has addressed directly the role of the prefrontal cortex in goal-directed instrumental action, if hypotheses derived from humans and primates are correct, the substantial anatomical similarity between the rat, primate and human prefrontal cortex suggests that instrumental learning should depend upon the integrity of this structure in the rat. In rats, the dorsolateral prefrontal cortex corresponds roughly to the prelimbic area (Fuster, 1989) and, as a consequence, we examined whether the prelimbic area (PL) of the prefrontal cortex plays a role in contingency learning by this animal. The behavioural effects of lesions of the PL area were contrasted with equivalent lesions of a second cortical area, the insular cortex (IC).

The reason for focusing on the IC arises from its role in gustatory processing. Our recent studies of reward devaluation by specific satiety suggest that the incentive value of a food reward can be carried by its taste alone (Balleine and Dickinson, 1998). Following instrumental training with two starch-based food rewards that differed only in their taste, prefeeding one of these foods prior to an extinction test selectively reduced performance of the response trained with that reward. The implication of this finding is that the incentive values of the two foods were associated with their tastes. Given that the assignment of incentive value is learned, the brain structures most likely to mediate this function are those involved in gustatory learning and, for this reason, we considered the IC to be a strong candidate for the site of the incentive memory on the basis of its established role in general gustatory learning (Braun, 1990; Braun et al., 1982; Rosenblum et al., 1997).

Within this framework, our research has contrasted the effects of neurotoxic lesions of PL and IC[2] on the sensitivity of instrumental performance to the response-reward contingency and to reward devaluation by specific satiety. Following recovery from the lesions, hungry rats were trained to press a lever and pull a chain for the food pellets and starch solution in separate sessions, with the response-reward assignments counterbalanced across animals. The lesions did not affect this initial acquisition, and the pretraining continued until all animals were reliably pressing the lever and pulling the chain on the schedule in which the first response in each second was rewarded with a probability of 0.05. The effect of reward devaluation by specific satiety was then assessed. As in the previous specific satiety study, half of the rats in each group were prefed the food pellets and the remainder the starch solution for 1 h prior to being given a choice between the two actions in an extinction test.

Following the reward devaluation test, lever pressing and chain pulling were re-established with the appropriate rewards, again in separate sessions and under the same probabilistic reinforcement schedule as that employed during the initial training. The sensitivity of instrumental performance to degradation of the instrumental contingency was then assessed. As in the previous contingency study, the first lever press or chain pull in each second continued to be paired with the delivery of the appropriate reward with a probability of 0.05. In addition, however, one type of reward, either the food pellet or the starch solution, was delivered with a probability of 0.05 following each second without an action so that responding had no effect on the probabil-

[2] The coordinates were for the PL lesions: AP: $+3.2$; L: ± 0.8; V: -3.0 and -4.0; and for the IC lesions: AP: $+1.2$; L: ± 3.5; V: -6.0 @ 20° to the vertical. In both cases 1 μl of 0.09 M quinolinic acid was injected bilaterally to produce the lesions.

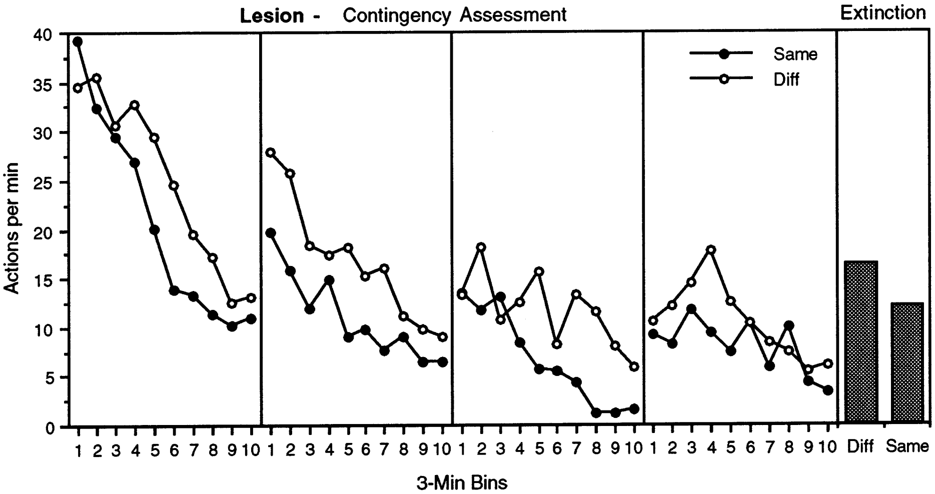

Fig. 4. The mean number of actions (lever pressing and chain pulling) per min by rats with lesions of the prelimbic area of the prefrontal cortex during the four sessions of training under the non-contingent training for each action and during the final extinction test (right-hand panel). The response rates are shown separately for the actions paired with the reward that was same as and different from the unpaired reward.

ity of reward. As we have seen, the impact of this non-contingent schedule on the performance of intact rats depends upon whether these unpaired rewards are the same as those paired with the action. Fig. 1 presents the performance of the control, sham-lesioned animals from the PL lesion experiment and, to recap, illustrates that the unpaired rewards reduced the action paired with the same type of reward more than that paired with the different reward. We interpreted this finding as evidence that intact rats are sensitive to the contingency between an action and a particular reward.

6.1. Contingency sensitivity

Fig. 4 shows that a very different pattern of responding is observed following PL lesions. For these animals, the introduction of unpaired rewards produced a profound reduction in the performance of not only the action paired with the same reward but also that paired with the different reward. Although on average the animals performed the different action slightly more than the same action, this discrepancy was very much smaller than for the sham-lesioned, control animals and was not reliable (see Fig. 1). That this insensitivity to the instrumental contingency is not a general product of cortical dysfunction is evident from the very different performance exhibited by animals with IC lesions. Fig. 5 shows a clear discrimination between the same and different actions following IC lesions that is comparable to the performance the control rats (see Fig. 1). On the final session of non-contingent training, all eight animals with IC lesions performed the action paired with the different reward more than that paired with the same reward.

There are, of course, many possible reasons why animals with PL lesions fail to show sensitivity to reward-specific contingencies. For example, they may be unable to discriminate between the two actions or between the two rewards. We think this is unlikely, however. Another group of rats with comparable PL lesions were trained on a discrimination in which the type of reward delivered in that session, the food pellets or the starch solution, signalled which of the two actions, lever pressing and chain pulling, was reinforced in any given session. The lesioned animals learned this discrimination as rapidly as controls, a feat that required the animals to distinguish between both the rewards and the actions. For this reason, we attribute the deficit on the non-contingent schedule to an inability to learn specific response-reward relationships with the consequence that the acquisition of the actions is primarily controlled by the S–R/reinforcement mechanism. The presentation of the unpaired rewards therefore produces a general disruption of performance, possibly through the conditioning of competing responses (see above).

6.2. Reward devaluation

If the performance of animals with the PL lesions is primarily controlled by an S–R mechanism, their performance should be unaffected by reward devaluation. In the absence of knowledge of the response-reward relationship, devaluing one of the rewards by prefeeding cannot selectively affect performance of the action trained with that reward. Fig. 6 displays the results of the specific satiety test for the PL animals and their sham-lesioned controls. The left panels show that the lesion had no effect on responding during the final training session and, moreover, that both lesioned and sham animals performed the action associated with the to-be-devalued reward at a comparable rate to that of

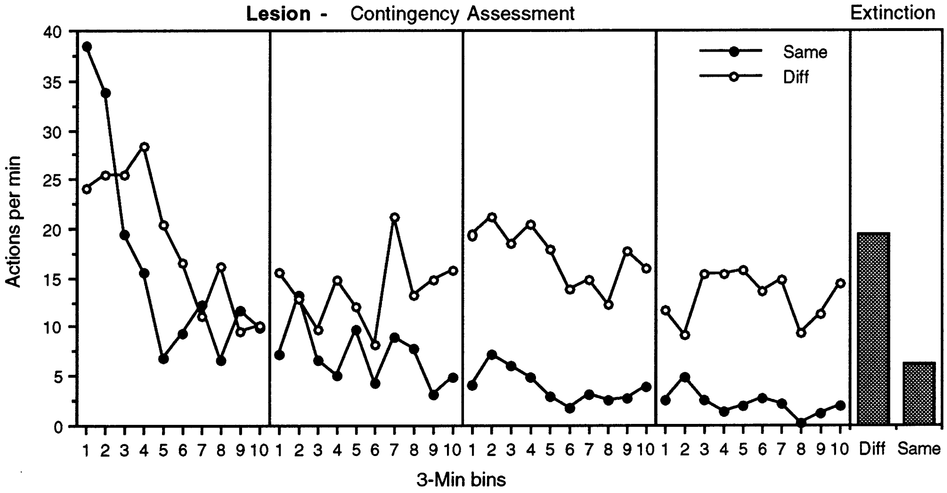

Fig. 5. The mean number of actions (lever pressing and chain pulling) per min by rats with lesions of the insular cortex during the four sessions of training under the non-contingent training for each action and during the final extinction test (right-hand panel). The response rates are shown separately for the actions paired with the reward that was same as and different from the unpaired reward.

the response trained with the reward which maintained its valued during the specific satiety test. More important, however, is the performance during the extinction test following prefeeding. The middle panel of Fig. 6 displays the strong devaluation effect observed in the sham control animals. From the outset of the test, all of the eight control rats performed the action associated with the non-prefed, and thus valued reward more than the one trained with the reward devalued by prefeeding. No such difference was observed in the lesioned animals, however. As the right-hand panel of Fig. 6 illustrates, these animals performed both actions at a low and indiscriminate rate from the outset of the test. Taken together with the results of the contingency study, the absence of a reward devaluation effect in rats with PL lesions suggests that they are essentially S–R, habit-driven animals with little knowledge of the consequences of their behaviour.

A very different interpretation must be given to the impact of IC lesions on reward devaluation. Fig. 7 shows that the profile of performance during the specific satiety test was very similar to that observed with PL dysfunction. Although the IC lesion also had no effect on the terminal level of instrumental performance on the last training session, it did abolish the effect of prefeeding during the extinction test. Whereas prefeeding selectively reduced performance of the action trained with prefed food in all of the eight sham control animals, the IC rats responded indiscriminately during the extinction test. In contrast to the PL lesion, however, the absence of a reward devaluation effect cannot be attributed to a lack of knowledge of the instrumental relationship for the animals with IC dysfunction were as sensitive to the reward-specific contingency as controls (Fig. 5).

In an attempt to determine the source of the IC deficit, the animals were tested in a second reward devaluation test. Performance was re-established on the probabilistic reinforcement schedule in two further training sessions for each action before the animals were again prefed one of the food rewards for 1 h. This prefeeding was immediately followed by a choice test but, unlike the first test, both responses were reinforced on the training schedule with the appropriate reward during the test. As Fig. 8 shows, the presence of the rewards during the test had a major effect on the sensitivity of the lesioned animals to prefeeding. Whereas these rats showed indiscriminate responding during the preceding extinction test (Fig. 7), the prefeeding selectively reduced performance of the action reinforced with the prefed reward as much in the IC lesioned animals as in the sham controls when the rewards were actually delivered.

This result demonstrates that an intact IC is not required for assigning a new incentive value to a food reward on the basis of the animal's current state when it is allowed direct experience with the reward. What is impaired, however, is the ability to store and retain information about the changed value with the consequence that this new value can not control performance through knowledge of the specific response-reward relationship in the absence of direct exposure the revalued reward. It is this ability that is challenged by the extinction test (Fig. 7).

6.3. Incentive learning

In our discussion of the motivational control of instrumental action, we presented evidence that such control is not always direct but that animals have to learn about the values of rewards in different motiva-

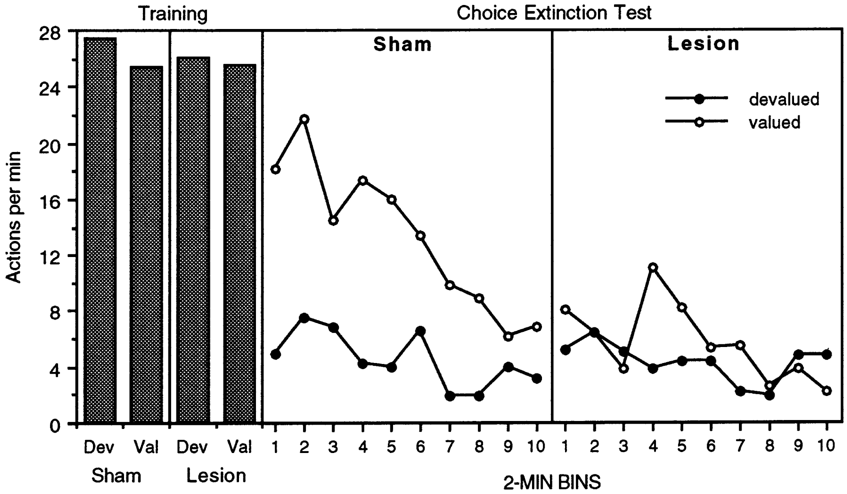

Fig. 6. The mean number of actions (lever pressing and chain pulling) per min on the last session of instrumental training (left-hand panels) and during the extinction test following prefeeding with one of the rewards (right-hand panels). The response rates are shown separately for the action trained with the reward that was devalued (Dev) by prefeeding and for the action trained with the reward that was not prefed and therefore should have retained its value (Val). The performance is shown separately for rats with lesions of the prelimbic area of the prefrontal cortex and a sham treatment.

tional states. For this reason, we argued that the incentive value of a food reward is not necessarily assigned just to a representation of the food alone but, under some circumstances, to a conjoint representation of the food and the hunger or deprivational state in which it has been experienced. It is this conjoint representation that allows motivational control over the performance of a goal-directed, instrumental action. A critical prediction from this analysis is that IC lesions should interfere with the motivational control by a hunger state over food-rewarded behaviour if this structure functions as the incentive memory for this class of rewards.

To assess this prediction, we conducted an incentive learning study with further groups of IC lesioned and sham rats. Initially, they were trained to lever press and chain pull for the food pellets and starch solution while hungry in exactly the same manner as in the previous studies. Following this training, they were given the opportunity to learn about the relative values of these rewards in the deprived or hungry state and in the non-deprived or sated state. During this incentive learning stage, the rats were given a series of six daily sessions in the feeding cages with their motivational state alternated daily between hunger and general satiety. The state of hunger was induced by depriving the animals of their maintenance diet in their home cages for the preceding 22.5 h, whereas they had free access to this diet during this period to induce a state of general satiety. Half of the animals in each group received the food pellets on days when they were sated and the starch solution on the days when they were hungry. This incentive learning experience should have allowed the animals to learn that the food pellets have a low value when they are in the sated state without the opportunity for learning about the reduced value of the starch solution in this state. The remaining animals received the opposite reward-motivational state assignment with the corresponding consequences for incentive learning. Finally, all animals were given free access to their maintenance diet and then given a choice between the actions during an extinction test in the sated state.

The left panel of Fig. 9 illustrates the incentive learning effect observed in the sham-lesioned, control animals. All but one of the eight rats in this group performed the action trained with the reward reexposed in the sated state during incentive learning less than that associated with the reward reexposed when the animals were hungry. By contrast, the relative performance of the two actions was similar for the animals with the IC lesions. During the incentive learning stage, the lesioned animals failed to store information about how variations in motivational state modulates the incentive value of a food reward in a way that transfers to the control of subsequent instrumental performance. As a consequence, these rats failed to show the appropriate reduction in responding when they were stated at the time of testing. This is just the deficit in incentive learning that we should expect if the IC is a component of the incentive memory for food rewards.

7. Summary and conclusions

The control of instrumental action is complex involving the interaction of at least three psychological processes. Although the insensitivity of performance to reward devaluation after overtraining implies a role for

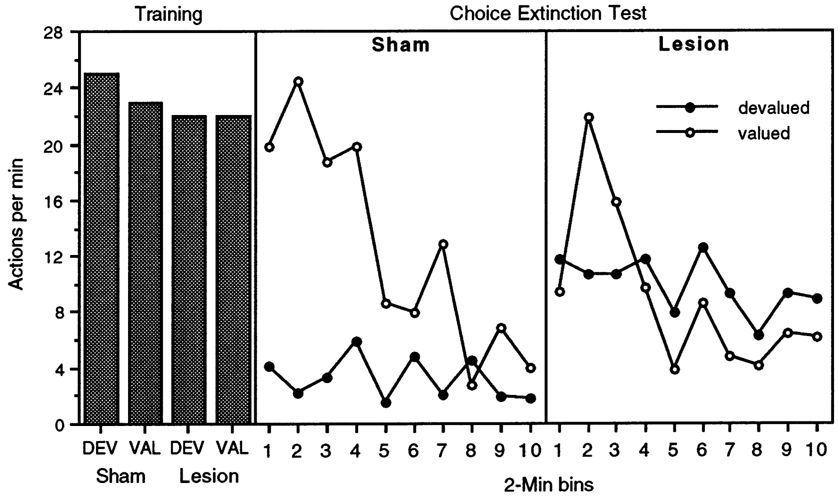

Fig. 7. The mean number of actions (lever pressing and chain pulling) per min on the last session of instrumental training (left-hand panels) and during the extinction test following prefeeding with one of the rewards (right-hand panels). The response rates are shown separately for the action trained with the reward that was devalued (DEV) by prefeeding and for the action trained with the reward that was not prefed and therefore should have retained its value (VAL). The performance is shown separately for rats with lesions of the insular cortex and a sham treatment.

the classic S–R/reinforcement mechanism, the fact that devaluing the reward after more limited training reduces responding in a subsequent extinction test demonstrates that instrumental action can be goal-directed. Goal-directed control is mediated by two further processes, contingency learning and incentive learning.

Contingency learning involves the acquisition of information about the relationship between the instrumental action and the reward and is sensitive not just the contiguity between response and reward but also to their contingency or causal association. Degrading the instrumental contingency by the presentation of unpaired rewards decreases responding when paired and unpaired rewards are the same. Incentive learning, on the other hand, mediates the effects of motivation variables on performance by controlling the acquisition of incentive value by the reward. Primary motivational states, such as hunger, do not necessarily have a direct impact on instrumental performance and animals have to learn about the modulation of incentive value produced by variations in such states.

Preliminary investigations of the brain structures involved in goal-directed instrumental action suggest that the PL and the IC function as components of the

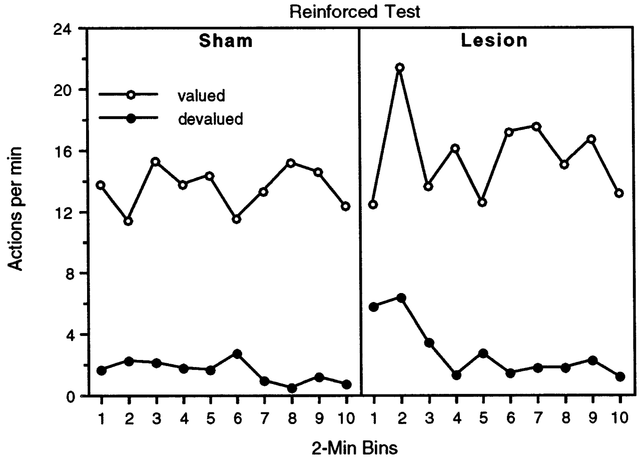

Fig. 8. The mean number of actions (lever pressing and chain pulling) per min during a reinforced test following prefeeding with one of the rewards. The response rates are shown separately for the action trained and tested with the reward that was devalued by prefeeding and for the action trained and tested with the reward that was not prefed and therefore should have retained its value. The right-hand panel illustrates the performance of animals with lesions of the insular cortex and the left-hand panel the performance following a sham treatment.

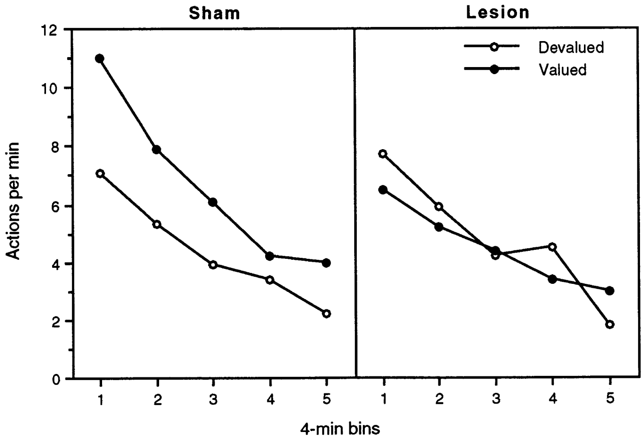

Fig. 9. The mean number of actions (lever pressing and chain pulling) per min during an extinction test while the rats were sated by free access to the maintenance diet. The response rates are shown separately for the action trained with the reward that was devalued by prior exposure in the sated state and for the action trained with the reward that was exposed in the hungry state and therefore should have retained its value. The right-hand panel illustrates the performance of animals with lesions of the insular cortex and the left-hand panel the performance following a sham treatment.

contingency and incentive memories, respectively. Lesions of the PL render rats insensitive to variations in the contingency between a response and a specific food reward suggesting that the instrumental behaviour of these animals is primarily habit based. By contrast, rats with IC dysfunction appear to encode the instrumental contingency but have deficits in responding to motivational manipulations which normally change the incentive value of a food reward.

Finally, it is worth noting that the tripartite account of instrumental learning and the role of cortical structures in the control of goal-directed action are compatible with an evolutionary trajectory for the genesis of instrumental behavior. At a psychological level, the contingency and incentive learning processes controlling goal-directed action are built upon and supplement the more simple mechanisms for habitual behaviour. Correspondingly, at the neurobiological level, it is the evolutionary more archaic subcortical structures that support habitual responding and provide the foundations upon which cortical representations of event relationships and values support the capacity for purposive and intentional action.

Acknowledgements

The research reported in this article and preparation of the manuscript was supported by grants from the National Institute of Mental Health, NIMH grant no. MH 56446, and from the European Commission BIOMED 2 programme.

References

Adams, C.D., 1982. Variations in the sensitivity of instrumental responding to reinforcer devaluation. Q. J. Exp. Psychol. 34B, 77–98.

Adams, C.D., Dickinson, A., 1981. Instrumental responding following reinforcer devaluation. Q. J. Exp. Psychol. 33B, 109–122.

Balleine, B.W., 1992. The role of incentive learning in instrumental performance following shifts in primary motivation. J. Exp. Psychol: Anim. Behav. Proc. 18, 236–250.

Balleine, B., Dickinson, A., 1998. The role of incentive learning in instrumental outcome revaluation by sensory-specific satiety. Anim. Learn. Behav. 26, 46–59.

Braun J.J., 1990. Gustatory cortex: Definition and function. In: Kolb B., Tees R., (Eds.), The Cerebral Cortex of the Rat. MIT Press, Cambridge, MA., pp. 407–430.

Braun, J.J., Lasiter, P.S., Kiefer, S.W., 1982. The gustatory neocortex of the rat. Physiol. Psychol. 10, 13–45.

Carp, J.S., Wolpaw, J.R., 1994. Motoneuron plasticity underlying operantly conditioned decrease in primate H-reflex. J. Neurophysiol. 72, 431–442.

Chen, X.Y., Wolpaw, J.R., 1995. Operant conditioning of H-reflex in freely moving rats. J. Neurophysiol. 73, 411–415.

Colwill, R.M., Rescorla, R.A., 1985a. Postconditioning devaluation of a reinforcer affects instrumental responding. J. Exp. Psychol: Anim. Behav. Proc. 11, 120–132.

Colwill, R.M., Rescorla, R.A., 1985b. Instrumental responding remains sensitive to reinforcer devaluation after extensive training. J. Exp. Psychol: Anim. Behav. Proc. 11, 520–536.

Colwill R.M. and Rescorla R.A. (1986) Associative structures in instrumental learning. In: Bower G.H., (Ed.) The Psychology of Learning and Motivation, 20: Academic Press, Orlando, FL., pp.55–104.

Colwill, R.M., Rescorla, R.A., 1988. The role of response-reinforcer associations increases throughout extended instrumental training. Anim. Learn. Behav. 16, 105–111.

Davidson, T.L., 1993. The nature and function of interoceptive signals to feed: Towards intergration of physiological and learning perspectives. Psychol. Rev. 100, 637–640.

Dickinson A., 1980. Contemporary Animal Learning Theory. Cambridge University Press, Cambridge.

Dickinson, A., Balleine, B., 1994. Motivational control of goal-directed action. Anim. Learn. Behav. 22, 1–18.

Dickinson, A., Balleine, B., 1995. Motivational control of instrumental action. Curr. Dir. Psychol. Sci. 4, 162–167.

Dickinson, A., Mulatero, C.W., 1989. Reinforcer specificity of the suppression of instrumental performance on a non-contingent schedule. Behav. Proc. 19, 167–180.

Dickinson A., Shanks D.R., 1995. Instrumental action and causal representation. In: Sperber D., Premack A.J., (Eds.) Causal Cognition: A Multidisciplinary Debate. Clarendon Press, Oxford, pp. 5–25.

Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., Boakes, R.A., 1995. Motivational control after extended instrumental training. Anim. Learn. Behav. 23, 197–206.

Dickinson, A., Campos, J., Varga, Z.L., Balleine, B.W., 1996. Bidirectional instrumental conditioning. Q. J. Exp. Pychol. 49B, 289–306.

Donahoe, J.W., Burgos, J.E., Palmer, D.C., 1993. A selectionist approach to reinforcement. J. Exp. Anal. Behav. 60, 17–40.

Donahoe, J.W., Palmer, D.C., Burgos, J.E., 1997. The unit of selection: What do reinforcers reinforcer. J. Exp. Anal. Behav. 68, 259–273.

Frith, C.D., Friston, K., Liddle, P.F., Frackowiak, R.S., 1991. Willed action and the prefrontal cortex in man: a study with PET. Proc. R. Soc. Lond. 244, 241–246.

Fuster J., 1989. The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe (2nd ed). Raven Press, New York.

Gallistel C.R., 1990. The Organization of Learning. MIT Press, Cambridge, MA.

Gluck, M.A., Thompson, R.I., 1987. Modeling the neural substrate of associative learning and memory: a computational approach. Psychol. Rev. 94, 176–191.

Goldman-Rakic, P.S., 1994. The issue of memory in the study of prefrontal function. In: Thiery, A.-M., et al. (Eds.), Motor and Cognitive Functions of the Prefrontal Cortex. Springer-Verlag, Berlin, pp. 112–121.

Grand S., Cliff D., Malhotra A., 1996. Creatures: Artficial Life Autonomous Software Agents for Home Entertainment. University of Sussex Technical Report CSRP434.

Hammond, L.J., 1980. The effects of contingencies upon appetitive conditioning of free-operant behavior. J. Exp. Anal. Behav. 34, 297–304.

Hawkins, R.D., Kandel, E.R., 1984. Is there a cell-biolological alphabet for simple forms of learning? Psychol. Rev. 91, 375–391.

Hetherington M.M., Rolls B.J., 1996. Sensory-specific satiety: Theoretical issues and central characteristics. In: Capaldi, E.D., (Ed.), Why We Eat What We Eat , American Psychological Association, Washington, D.C., pp. 267–290.

Hershberger, W.A., 1986. An approach through the looking glass. Anim. Learn. Behav. 14, 443–451.

Hollis, K.L., Pharr, V.L., Dumas, M.J., Britton, G.B., Field, J., 1997. Classical conditioning provides paternity advantage for territorial male blue gouramis (*Trichogaster tricopterus*). J. Comp. Psychol. 111, 219–225.

Hull C.L., 1943. Principles of Behavior. Appleton-Century-Crofts, New York.

Knight, R.T., Grabowecky, M.F., Scabini, D., 1995. Role of human prefrontal cortex in attention control. Adv. Neurol. 66, 21–34.

Lavond, D.G., Kim, J.J., Thompson, R.F., 1993. Mammalian brain substrates of aversive classical conditioning. Ann. Rev. Psychol. 44, 317–342.

Ledoux J., 1995. The Emotional Brain. Simon & Schuster, New York.

McGaugh J.L., Weinberger N.M., Lynch G., 1995. Brain and Memory: Modulation and Mediation of Neuroplasticity. Oxford University Press, New York.

Pearce, J.M., Hall, G., 1980. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol. Rev. 87, 532–552.

Rescorla R.A., Wagner A.R., 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In: Black A.H., Prokasy W.F. (Eds.), Classical Conditioning II: Current Research and Theory, Appleton-Century-Crofts, New York, pp. 64–99.

Robbins, T.W., Everitt, B.J., 1996. Neurobehavioural mechanisms of reward and motivation. Cur. Opin. Neurobiol. 6, 228–236.

Rosenblum, K., Berman, D.E., Hazvi, S., Lamprecht, R., Dudai, Y., 1997. NMDA receptor and the tyrosine phosphorylation of its 2B subunit in taste learning in the rat insular cortex. J. Neurosci. 17, 5129–5135.

Schmajuk N.A., 1997. Animal Learning and Cognition. Cambridge University Press, Cambridge.

Squire, L.R., 1992. Memory and the hippocampus: A synthesis from findings with rats, monkeys and humans. Psychol. Rev. 99, 195–231.

Thorndike E.L., 1911. Animal Intelligence: Experimental Studies. Macmillan, New York.

Tolman E.C., 1932. Purposive Behavior in Animals and Man. Century, New York.

Wagner A.R., 1981. SOP: A model of automatic memory processing in animal behavior. In: Spear, N.E., Miller R.R. (Eds.), Information Processing in Animals: Memory Mechanisms, Lawrence Erlbaum Associates, Hillsdale, N.J., pp. 5–47.

White, N.M., 1989. A functional hypothesis concerning the striatal matrix and patches: Mediation of S–R memory and reward. Life Sci. 45, 1943–1957.

Williams, B.A., 1989. The effect of response contingency and reinforcement identity on response suppression by alternative reinforcement. Learn. Motiv. 20, 204–224.

Wolpaw, J.R., Lee, C.L., Calaitges, J.G., 1989. Operant conditioning of primate triceps surae H-reflex produces reflex asymmetry. Exp. Brain Res. 75, 35–39.