1.36 Theory of Reward Systems

S. B. Ostlund, N. E. Winterbauer, and B. W. Balleine, University of California, Los Angeles, CA, USA
© 2008 Elsevier Ltd. All rights reserved.

1.36.1 Introduction

Except, perhaps, for the unabashed moralist, knowledge alone does not determine choice of action, i.e., knowing that "action A leads to X and action B leads to Y" does not ‘entail’ choosing A or B. What enables choice, given this information, is some nonarbitrary means of establishing the relative merits of achieving X or Y. We argue that it is the reward system that provides this means. From this perspective, although the reward system is an extension of the general motivational processes of animals, its function is limited to actions over which animals can exert control and that are instrumental to achieving some goal or other, i.e., to goal-directed instrumental actions. Of course, although most aspects of an animal's behavioral repertoire can be described in goal-directed terms, many of these activities are not goal-directed at all and are reflexive responses elicited by stimuli or relations between stimuli. Establishing criteria for discerning goal-directed and non-goal-directed actions is a necessary step, therefore, in limiting our discussion of the reward system. In this chapter, we consider first the criteria for defining an action as goal-directed and then use that definition to describe the nature and function of the reward system in establishing primary rewarding events, like foods and fluids, both with respect to encoding reward value and to retrieving that value in order to choose between competing courses of action.

This research has established that the value of reward is determined by the quality of the emotional response associated with an event, the latter dependent on current motivational state, i.e., value essentially maps onto the relationship between the specific sensory features of an event and the particular, pleasant or unpleasant, emotional feedback generated when that event is contacted. This issue is taken up in more detail in the section titled ‘Reward processes,’ where we examine one of the main predictions of this account, that, in the context of secondary rewards, any event associated with a pleasant emotional reaction will support the performance of goal-directed actions. These are sensory events that acquire reward value through association with primary rewards (commonly mislabeled conditioned reinforcers). The procedures used to establish secondary rewards are identical to those commonly used to establish Pavlovian conditioned responses to a stimulus, raising the possibility that the functioning of the reward system can be reduced to the motivational processes that support Pavlovian conditioning. In the section titled ‘Secondary reward’ we examine this possibility and conclude, based on the extensive evidence standing against this claim, that Pavlovian conditioned responses (CRs) and
goal-directed actions are controlled by fundamentally distinct incentive processes.

1.36.2 Reward Processes

1.36.2.1 Goal-Directed Actions and Behavioral Control

The critical distinction between reflexive and goal-directed actions is that the latter are controlled by a causal relationship to their consequences, whereas the former are not. There are many illustrations of this distinction but perhaps the most apposite is Sheffield’s (1965) analysis based on the salivary response of dogs. Salivation was the conditioned and unconditioned reflex studied by Pavlov (1927). Nevertheless, from a goal-directed perspective, it is possible that dogs control this response in order to facilitate digestion or to improve the taste of food. Sheffield arranged a standard pairing between conditioned and unconditioned stimuli, in this case presentation of a tone followed by food delivery, but with a twist: If the dog salivated during the tone the food was not delivered on that trial. This arrangement maintains a Pavlovian relationship between the tone and food but abolishes any instrumental contingency between salivation and food. He reasoned that, if the salivation was goal-directed then this omission contingency should ensure that they stop salivating; indeed having never had the opportunity to learn that salivating improved the rewarding impact of the food by enhancing its flavor or improving its ingestion, they should never acquire salivation to the tone in the first place. Sheffield found that it was clearly the Pavlovian relationship controlling performance; during the course of over 800 tone-food pairings the dogs acquired and maintained salivation to the tone even though this resulted in them losing most of the food they could otherwise have obtained.

Salivation may be the exception of course, but in numerous studies over the last 40 years it has been established in a range of species that Pavlovian conditioned responses do not adjust to this kind of contingency, i.e., one in which performance of the conditioned response leads to the omission of the unconditioned stimulus. Rats acquire conditioned approach responses during a conditioned stimulus (CS) when doing so omits the food (Holland, 1979), pigeons peck at keys (Williams and Williams, 1969), chicks chase food away (Hershberger, 1986), and so on. In all of these studies, the evidence confirms that the performance of the Pavlovian CR does not depend on the relationship between the CR and the US. (See Chapter 1.03).

In contrast, experiments assessing the performance of actions acquired during instrumental conditioning have found evidence that these responses do indeed depend on the contingency between action and outcome. Take, for example, instrumental lever pressing. Rats will acquire lever pressing for food quickly and without explicit shaping. Putting this response on an omission contingency, in which responding leads to the omission of an otherwise freely delivered food, rapidly reduces the performance of that response, more rapidly than simply delivering the outcome in an unpaired manner (Davis and Bitterman, 1971; Dickinson et al., 1998; See Chapter 1.06). Furthermore, numerous studies have demonstrated the exquisite sensitivity of the performance of instrumental lever pressing to changes in the net probability of outcome delivery given the action (i.e., the difference between probability of an outcome given a response and the probability of the outcome given no response). These changes can be highly selective; degrading one action–outcome contingency by delivering the outcome associated with that action noncontingently often has no effect on the performance of other actions (Colwill and Rescorla, 1986; Dickinson and Mulatero, 1989; Balleine and Dickinson, 1998a).

1.36.2.2 The Effect of Changes in Reward Value

Generally, therefore, goal-directed actions are those that, unlike Pavlovian CRs, are sensitive to the causal relation between the performance of the action and its specific outcome. It is important to note, however, that lever press responses can be controlled by two kinds of association. The first is the relationship between action and outcome described earlier. After extensive instrumental training, however, performance of an action can become habitual, elicited by various situational cues connected with the action through a process of sensorimotor association (Adams, 1981; Dickinson, 1985, 1994). Although the formation of these associations diminishes sensitivity to omission (Dickinson et al., 1998), it does not necessarily abolish it and, although this test distinguishes actions from Pavlovian conditioned reflexes, it does not provide an adequate assessment in itself to distinguish goal-directed actions from habits. Fortunately, there is a clear distinction between the functions of the instrumental outcome in the two
forms of learning. Whereas the outcome serves as the second term of the action–outcome association that supports the acquisition and performance of goal-directed actions, it serves merely to strengthen or to reinforce the stimulus–response (S–R) associations that form habits. As such, the outcome forms no part of the associative structure that supports habitual performance. Based on this analysis, therefore, and combined with an assessment of sensitivity to changes in the instrumental contingency, the standard test of whether an action is goal-directed or not involves an assessment of the sensitivity of performance to a posttraining change in the reward value of the outcome. From an S–R perspective, when conducted posttraining, i.e., after a substantial S–R association has been established, a change in outcome value should be expected to have little if any effect on the subsequent tendency to perform the action. If an action is goal-directed, however, the change in value should potently alter performance.

Consider the case in which a hungry rat is trained to press a lever for a particular type of food pellet. According to a goal-directed account, it is the reward value of the food pellets that motivates performance. Consequently, if having trained the rat to perform this action, the reward value of the food pellets is reduced in some way, we should expect this devaluation to affect performance, i.e., the rat should be less inclined to press the lever after the devaluation. Given this scenario, the question at issue is whether the devaluation affects performance via the animal’s knowledge of the contingency between lever pressing and the food pellets. In the first appropriately controlled study along these lines, Adams and Dickinson (1981) assessed this by training rats with two types of food pellets, sugar and grain, with only one type being delivered by lever pressing. The other type of pellet was presented independently of any instrumental action. Thus, any particular rat might have to work for sugar pellets by lever pressing, while receiving free deliveries of grain pellets every so often. The issue was whether the animals would reduce lever pressing more after the devaluation of the response-contingent pellets, the sugar pellets in our example, than after devaluation of the free pellets, the grain ones. Such an outcome could only occur if the effect of the devaluation was mediated by the instrumental contingency between lever pressing and the sugar pellets.

In this study, the pellets were devalued using conditioned taste aversion procedures; it is well established that a food aversion can be conditioned by inducing gastric illness, for example by the injection of lithium chloride (LiCl), shortly after the animal has consumed the food (Bernstein, 1999). In the Adams and Dickinson study, having trained the rats to lever press, half had a taste aversion conditioned to the sugar and half to the grain pellets. During aversion conditioning, the levers were withdrawn and the animals were given a series of sessions in each of which they were allowed to eat one type of pellet. The animals in the devaluation group received a LiCl injection after sessions in which they received the pellets that had been contingent on lever pressing during training but not following sessions with the free pellets. The control group, by contrast, had the aversion conditioned to the free pellets rather than the response-contingent ones. Although such food aversions can be established with a single pairing of consumption with illness when the food is novel, the treatment had to be repeated a number of times to suppress consumption in the present study. This is because the pellets were already familiar to the rats, having been presented during instrumental training.

After inducing these aversions, Adams and Dickinson were now in a position to ask whether devaluing the pellets that acted as the reward for lever pressing during training had a greater impact on performance than devaluing the freely delivered pellets. This result would be expected if the motivational properties of rewards are mediated by their instrumental relation to the action. In fact, this is just what Adams and Dickinson found: when subsequently given access to the lever again, the devaluation group pressed significantly less than the control group. Note that this test was conducted in extinction, during which neither type of pellet was presented, for if the pellets had been presented during testing, the reluctance of the devaluation group to press the lever could be explained simply in terms of the direct suppressive effect of presenting this aversive consequence. By testing in extinction, however, different performance in the two groups must have reflected integration of knowledge of the consequences of lever pressing acquired during training with the current reward value of the pellets. This suggestion was further confirmed by Colwill and Rescorla (1986) using a choice test. They trained hungry rats to perform two instrumental actions, lever pressing and chain pulling, with one action earning access to food pellets and the other earning access to a sucrose solution. The rats were then given several trials in which they were allowed to consume one of the outcomes with the levers and chains.
withdrawn and were then made ill by an injection of LiCl. All animals were then given a choice extinction test on the levers and chains again conducted in extinction, i.e., in the absence of either of the outcomes. Although S–R accounts should predict no effect of this treatment, Colwill and Rescorla found that animals performed less of the action whose training outcome was subsequently paired with LiCl than the other action, indicating that the rats had indeed encoded the consequences of their actions.

The importance of these demonstrations of the outcome devaluation effect lies in the fact that, together, they provide strong evidence that animals encode the specific features of the consequences or outcome of their instrumental actions. Furthermore, these studies show that instrumental performance is not only determined by the encoding of the action–outcome relation but also by the current reward value of the outcome. In recent years, considerable attention has been paid to the processes that contribute to the encoding of reward value, and the advances that have been made have come largely from asking how outcome devaluation works to change instrumental performance: How does taste aversion work to modify the rats’ evaluation of the outcome and so change the course of its instrumental performance?

1.36.2.3 Incentive Learning and the Encoding of Reward Value

Perhaps the simplest account of the way taste aversion learning works to devalue the instrumental outcome can be derived from accounts of aversive conditioning generally according to which pairing the instrumental outcome with illness changes the evaluation of the outcome through the formation of a predictive association between the food or fluid and the aversive state induced by illness. The result of an effective pairing of the outcome with illness is, therefore, that the animal learns that the outcome now signals that aversive consequence. From this perspective, the outcome devaluation effect is the product of a practical inference process through which a previously encoded action–outcome relation is combined with learning that the outcome signals an aversive consequence to reduce subsequent performance of the action.

In contrast, Garcia (1989) introduced a more complex account according to which the change in the evaluation of the outcome induced by taste aversion learning is not due to changing what the outcome predicts but due to changes in how it tastes. Garcia related the change in taste to negative feedback from a system sensitive to illness that he identified as inducing a disgust or distaste reaction. It is important to see that this view implies that taste aversion learning involves not one learning process but two: (1) an effective pairing of the outcome with illness initially enables a connection between the sensory properties of the outcome and processes sensitive to illness; (2) this association is activated when the outcome is subsequently contacted to generate a distaste reaction and allow the animal to associate the outcome representation with disgust or distaste. This account predicts that, to induce outcome devaluation, it is not sufficient merely to pair the outcome with an injection of LiCl. Rather, a change in value is not induced until the second process is engaged when the outcome is again contacted.

The procedures employed to induce instrumental outcome devaluation, such as that described by Adams and Dickinson (1981), do not differentiate between these two accounts of taste aversion learning because the conditioning of an aversion to the outcome is usually conducted using multiple pairings of the outcome with illness. Clearly the pairings themselves would be sufficient to establish a signaling relation between the outcome and an aversive consequence. But the fact that the animals were allowed to contact the outcome on subsequent pairings could have provided the opportunity for the animals to associate the outcome representation with distaste. If a substantial aversion to the outcome could be conditioned with a single pairing of the outcome with illness, however, then these accounts of outcome devaluation make divergent predictions: On the signaling account, a devaluation effect should emerge, providing that an effective pairing between the taste and illness was produced; on Garcia’s (1989) account it should not emerge until the rats have been reexposed to the devalued outcome. In a test of these divergent predictions, Balleine and Dickinson (1991) trained thirsty rats to lever press for water. After acquisition, the outcome was switched to sugar solution for a single session, after which the rats were given an injection of LiCl either immediately or after a delay (the latter treatment, as an unpaired control, should have induced relatively little aversion to the sucrose on either account). The critical question was whether, in the absence of further contact with the sucrose, the rats in the immediately poisoned group...
would display reduced performance on the lever relative to the delayed group.

To assess the influence of reexposure to the sucrose, half of each of the immediate and delayed groups were allowed merely to taste the sucrose, whereas the remainder were given water before two tests were conducted on the levers. The first test was conducted in extinction to assess the effects of devaluation and reexposure on the tendency to press. A second, punishment test was then conducted in which responding on the lever again delivered the sucrose, which allowed us to assess the strength of the aversion to sucrose. If a substantial aversion to the sucrose was conditioned in the immediately poisoned groups, then not only should a reliable punishment effect have emerged in the second test, but, on the signaling account, responding should also have been reduced in the extinction test in all of the immediately poisoned rats. In contrast, in Garcia's account, responding in the extinction test should be reduced in those immediately poisoned rats given reexposure to the sucrose. In fact, in this and in several other experiments along similar lines, Balleine and Dickinson (1991) and Balleine (1992) found consistent evidence for Garcia's account; although a single pairing between sucrose and illness invariably produced a reliable punishment effect in immediately poisoned rats, a devaluation effect only emerged in the critical extinction test if reexposure to the sucrose was given prior to the test.

These results suggest that outcome devaluation depends upon the interaction of two learning processes. The first process involves the conditioning of an association between the outcome and processes that are activated by the induction of illness by LiCl. The failure of this learning process to directly affect instrumental performance suggests that it is not, alone, sufficient to induce outcome devaluation. Rather, it appears to be necessary for feedback from this first learning process to become explicitly associated with the specific sensory features of the outcome itself for a change in the reward value of the instrumental outcome to occur and for performance to change. Indeed, considerable evidence now suggests that this second learning process critically determines the encoding of the rewarding properties of the instrumental outcome, a process referred to as incentive learning (Dickinson and Balleine, 1994, 1995).

The reason for emphasizing the role of incentive learning in instrumental outcome-devaluation effects is that it also appears to be the process by which other primary motivational states, such as hunger and thirst, encode the reward value of other goals such as foods and fluids. It is well established that the motivational state of rats is a major determinant of their instrumental performance; not surprisingly, hungry animals work more vigorously for a food reward than sated ones. But what current evidence suggests is that this is because a food-deprived state induces an animal to assign a higher incentive value to nutritive outcomes when they are contacted in that state and that this high rating of the incentive value of the outcome is then reflected in a more vigorous rate of performance. Although this suggestion stands contrary to general drive theories of motivation that suppose that increments in motivation elicit their effects on performance by increases in general activation (Hull, 1943), there are good empirical grounds for arguing that motivational states do not directly control performance (Dickinson and Balleine, 1994, 2002; Balleine, 2001). Balleine (1992) trained groups of undeprived rats to lever press for a food reward. After training, half of the rats were shifted to a food deprivation schedule, whereas the remainder were maintained undeprived before both groups were given an extinction test on the levers. Balleine found that performance of the groups on test did not differ even though the shift in motivational state was clearly effective. In a subsequent test where the animals could again earn the food pellets, the food-deprived rats pressed at a substantially higher rate than the undeprived rats. Although motivational state clearly did not exert any direct control over performance, as was found in taste aversion conditioning, the motivational state could control performance if the rats were given the opportunity for incentive learning by allowing them consummatory contact with the instrumental outcome in the test motivational state prior to the test. To demonstrate this, Balleine (1992) trained two further groups of rats to lever press when undeprived. Both groups were given prior exposure to the instrumental outcome when food-deprived before the test in which one group was tested undeprived and the other food-deprived. Now a clear difference in performance was found in that the rats tested when food-deprived and allowed to consume the instrumental outcome when food-deprived prior to test pressed at a higher rate than the other three groups that in turn did not differ. Balleine (1992) was able to confirm that this incentive learning effect depended upon the instrumental contingency. He trained undeprived rats to perform two actions, lever pressing and chain pulling,
with one action earning access to food pellets and the other to a maltodextrin solution. All rats were then given a choice extinction test on the levers and chains. Prior to the test, however, the animals were given six sessions in which they were allowed to consume one instrumental outcome when food deprived and, on alternate days, the other outcome in the training, i.e., undeprived, state. On test, Balleine found that animals performed more of the action that, in training, had delivered the outcome reexposed in the food-deprived state prior to the test than the other action.

It should be noted that this role for incentive learning in instrumental performance following a shift in motivational state is not confined to posttraining increases in food deprivation. The same pattern of results was also found for the opposite shift, i.e., where rats were trained to lever press for food pellets when food-deprived and then tested when undeprived. In this case, rats only reduced their performance when food deprivation was reduced if they were allowed to consume the instrumental outcome when undeprived prior to the test (Balleine, 1992; Balleine and Dickinson, 1994). Finally, the generality of this role of incentive learning in instrumental performance has been confirmed for a number of different motivational systems and in a number of devaluation paradigms. For example, in addition to taste aversion learning, incentive learning has been found to mediate (1) specific satiety-induced outcome devaluation effects (Balleine and Dickinson, 1998b); (2) shifts from water deprivation to satiety (Lopez et al., 1992); (3) changes in outcome value mediated by drug states (Balleine et al., 1994, 1995a); and changes in the value of (4) thermoregulatory rewards (Hendersen and Graham, 1979) and (5) sexual rewards (Everitt and Stacey, 1987; Woodson and Balleine, 2002) (see Dickinson and Balleine, 1994, 2002; Balleine, 2001, for reviews). In all of these cases, it is clear that animals have to learn about changes in the incentive value of an instrumental outcome through consummatory contact with that outcome before this change will affect performance.

1.36.2.4 Incentive Learning as an Emotional Process

Traditional neobehaviorist learning theories argued that CRs, what were called fractional anticipatory goal responses, could exert a motivating effect on instrumental performance (Hull, 1943, 1952; Spence, 1956). Largely due to the subsequent work of Konorski (1967) and Mowrer (1960), however, it is now widely accepted that these effects reflect the conditioning of an affective state that can exert a direct modulatory influence over consummatory responses and, through a change in the emotional responses elicited during ingestion, on instrumental performance (Rescorla and Solomon, 1967; Dickinson, 1989). Recent research investigating the microstructure of orofacial taste reactivity responses in rats to various tastes has provided evidence, not only of specific ingestion and rejection responses to sweet and bitter tastes, but also that the ingestive taste reactivity responses are increased in hungry rats to tastes previously paired with nutrients (Myers and Sclafani, 2001). Likewise, rejection-related taste reactivity responses are increased to tastes previously paired with illness (Berridge et al., 1981). With respect to incentive learning, this approach suggests that, during this form of consummatory exposure, activation of the outcome representation activates its associated motivational system, which, through activation of attendant affective processes, generates feedback in the form of an emotional response. This process is illustrated in Figure 1. On this account, incentive learning depends on two processes: a feedback process: (Figure 1 (a), (b)) and a feedforward process (Figure 1 (c)). Presenting the instrumental outcome in some motivational state or other provides the opportunity for the formation of an association between the outcome representation and the motivation system (Figure 1(a)) that acts to open a feedback loop (Figure 1(b)). When the outcome is subsequently contacted, activation of the outcome representation acts to produce specific motivational activity that results directly in activity in affective structures productive of an emotional response. Incentive learning

![Figure 1](image-url)

**Figure 1** The structure of incentive learning. (a) Sensory features of the instrumental outcome (Se) are associated with a motivational process (M). (b) Through connections with affective structures (A) this connection provides feedback in the form of an emotional response (Rem). (c) Incentive learning reflects the association between Se and Rem based on their contiguous activity.
(Figure 1(c)), then, is the formation of a feedforward association between the outcome representation and an emotional response.

Taste aversion-induced outcome devaluation effects provide a good example of this process. In this case, this perspective argues that a taste is first associated with activation of a disgust system induced by LiCl. After this pairing, reexposure to the taste can drive the disgust system to activate the aversive affective system to generate an aversive emotional response. It is the contiguous pairing of the taste and the emotional response that, from this perspective, drives the reduction in reward value induced by reexposure. Notice that, if pairing a taste with illness conditions an association between the taste and disgust, then blocking the activity of the disgust system at the time of conditioning using an antiemetic, i.e., a drug that prevents or relieves illness or nausea, should be predicted to attenuate the formation of that association with the effect that, in the test sessions, rats should prefer a taste poisoned under the antiemetic to some other poisoned taste. But furthermore, if the expression of a previously conditioned aversion, and the consequent change in reward value, depends upon the ability of the taste representation to access the disgust system via an established connection, blocking the activity of the disgust system with an antiemetic during reexposure should be predicted to block the incentive learning effect; see Figure 2.

In accord with this suggestion, Limebeer and Parker (2000) reported that the antiemetic ondansetron blocked the expression of the aversive taste reactivity responses induced by a taste previously paired with illness. Furthermore, we have assessed this prediction by assessing the influence of ondansetron on reexposure to a poisoned taste on instrumental choice performance (Balleine et al., 1995b). In this experiment, thirsty rats were trained in a single session to perform two actions, lever pressing and chain pulling, with one action delivering a sucrose solution and the other a saline solution on a concurrent schedule. Immediately after this training session, all of the rats were given an injection of LiCl. Over the next 2 days the rats were given brief periods of reexposure to both the sucrose and the saline solutions. Prior to one reexposure session, rats were injected with ondansetron in an attempt to block the emotional effects of reexposure, whereas prior to the other session they were injected with vehicle. The next day, the rats were given a choice extinction test on the lever and chain. If reexposure devalues the instrumental outcome via the ability of the outcome representation to access the disgust system, blocking the activity of that system with ondansetron should attenuate the effects of reexposure such that, on test, the action that, in training, delivered the outcome subsequently reexposed under ondansetron should be performed more than the other action. This is, in fact, exactly what was found (Balleine et al., 1995a). The attenuation of incentive learning by ondansetron provides, therefore, strong confirmation of the suggestion that incentive learning depends critically upon negative feedback generated by an association between the outcome representation and a disgust system.

### 1.36.2.5 Retrieving Reward Value

Given the role of incentive learning in the encoding of reward, it is interesting to consider how the value conferred by this process is retrieved to determine choice performance. Because the choice tests are often conducted many days after incentive learning, in extinction the rat is forced to rely on their memory of specific action–outcome associations and the current relative value of the instrumental outcomes. So how is value encoded for retrieval during this test?
A currently influential theory, the somatic marker hypothesis (Damasio, 1994), proposes that value is retrieved through the operation of the same processes through which it was encoded. According to this view, decisions based on the value of specific goals are determined by reexperiencing the emotional effects associated with contact with that goal. With regard to outcome devaluation effects, for example, the theory could not be more explicit:

When a bad outcome connected with a given response option comes to mind, however fleetingly, you experience an unpleasant gut feeling... that forces attention on the negative outcome to which the given action may lead, and functions as an automated alarm signal which says: Beware of danger ahead if you choose the option that leads to this outcome. The signal may lead you to reject, immediately, the negative course of action and thus make you choose between other alternatives (Damasio, 1994: 173).

An alternative theory proposes that reward values, once determined through incentive learning, are encoded abstractly (e.g., X is good or Y is bad and so on) and, as such, from this perspective they are not dependent on the original emotional effects induced by contact with the goal during the encoding of incentive value for their retrieval (see Balleine and Dickinson, 1998a; Balleine, 2005, for further discussion).

We have conducted several distinct series of experiments to test these two hypotheses and, in all of these, the data suggest that after incentive learning, incentive values are encoded abstractly and do not involve the original emotional processes that established those values during their retrieval (Balleine and Dickinson, 1994; Balleine et al., 1994, 1995a,b). One test of these two accounts was derived from consideration of the role of associations between the outcome representation and the disgust system in outcome devaluation described in the previous section. If the impact of outcome devaluation on performance is carried by emotional feedback induced by activation of the disgust system by the outcome representation, then, according to the somatic marker hypothesis, reducing the ability of the outcome representation to activate the disgust system during retrieval of incentive value on test by administering ondansetron prior to the test should be predicted to attenuate the effects of outcome devaluation on performance. This experiment replicated the procedures used in the experiment described earlier (Balleine et al., 1995b) except that, prior to the choice extinction test, half of the animals were injected with ondansetron, whereas the remainder were injected with vehicle. Based on the previous study, it was anticipated that the group given the injection of vehicle prior to the test would perform more of the action that, in training, had delivered the outcome reexposed under ondansetron. More importantly, if activation of the disgust system critically mediates the retrieval of incentive value during the test, as the somatic marker hypothesis suggests, then any difference found in the vehicle group should be attenuated in the group injected with ondansetron on test.

The results of this experiment were very clear; contrary to predictions of the somatic marker hypothesis, the injection of ondansetron on test had no impact whatsoever on performance in the choice extinction test. Whether injected with vehicle or ondansetron prior to the test, the action that, in training, delivered the outcome reexposed under ondansetron was performed more than the other action and to a similar degree. This finding suggests that, although activity in the disgust system determines the effects of incentive learning, the disgust system does not play a role once incentive learning has occurred, i.e., the retrieval of incentive value is not based on the same process through which it was encoded. In line with the proposal that reward value is encoded abstractly or symbolically and in contradiction to predictions from the somatic marker hypothesis position, in this and other similar studies we have found that the processes that determine the encoding of reward value are not required during the retrieval of that value during free choice tests in order for animals to select a course of action.

1.36.3 Secondary Reward

The suggestion that the reward process supporting instrumental conditioning is derived from an association between the sensory features of an event and an emotional response, together with the evidence for the abstract encoding of reward value, provides an immediate explanation as to how events not directly associated with primary motivational systems can serve as the goals of instrumental actions; from this perspective, any stimulus associated with an emotional response should be able to serve as a goal and so support the performance of goal-directed actions. In the past, the seemingly arbitrary nature of goals has been explained in terms of a process called
conditioned reinforcement (Skinner, 1938). Within that literature, this process was proposed as the means by which arbitrary things, like colored pieces of paper, could serve as reinforcers supporting the development of new response tendencies through the acquisition of various stimulus-response associations. It is our view that the term conditioned reinforcement is a misnomer; it implies that the actions that they support are no more than habits. Of course, most human actions are acquired and maintained by goals that are associated with primary rewards and so have only an indirect connection to primary motivational systems and as such are more likely to be goal-directed actions than habits. We propose that the process that determines the acquisition of these goals be referred to, therefore, as secondary reward (SdR). Nevertheless, it is clear that the goal-directed status of these actions is something that stands in need of direct assessment.

There is, in addition, a further implication of this account. Although one should anticipate that secondary rewards will be the more potent, what this account of incentive learning portends is that, if the emotional response associated with an event determines whether it can serve as a goal, essentially any event can serve as the goal of an action providing it induces a positive change in emotional tone. In this section we describe research indicating that both stimuli associated with already established rewards and salient sensory events can serve as goals, allowing animals to acquire new responses based on the relationship of actions to these, sometimes weakly but nevertheless apparently rewarding, consequences.

1.36.3.1 Sensory Versus Secondary Reward

As mentioned, the older literature dealing with the phenomenon of conditioned reinforcement proposed that when neutral stimuli were associated with reinforcing ones, they could become conditioned reinforcers. A problem widely neglected within this literature, however, is the fact that apparently neutral stimuli turn out to be very difficult to come by. Indeed, the vast majority of experimentally utilized stimuli are demonstrably not neutral with respect to their ability to support instrumental responding even prior to any pairing with primary reward (Kish, 1966). The capacity of environmental stimuli, or more correctly, of change in the state of environmental stimuli to support instrumental behavior can, however, be well enough handled by the current claim that reward value is controlled by the emotional response associated with that event providing it is accepted that change in environmental stimuli provides a sufficiently positive change in that response. From this perspective, therefore, events that are sufficiently mild to induce a positive change provide a source of sensory reward (SeR), whether it is derived from generalization or perhaps by another source of motivation, such as a form of preparatory state produced by general affective arousal (Konorski, 1967) or perhaps, as has occasionally been proposed in the past, by a primary motivational process such as curiosity (Berlyne, 1960).

In order to use secondary reward as a tool to establish the way apparently arbitrary events can become the goals of instrumental action, it is important first to compare the influence of secondary and sensory rewards on the performance of actions. The question is, which secondary reward procedure should one employ to do so? The central position of this notion in Hull’s conception of learning (Hull, 1943) and Skinner’s utilization of it to explain the origin of human actions without apparent reinforcement (Skinner, 1938) drove considerable research during the middle part of the last century intended to establish or to disprove its applicability to the conditioning process. The most commonly used procedure to analyze SdR has been in chain schedules of instrumental reinforcement, where both instrumental training with the SdR and the pairing of the event with reward presumed to support that conditioned reinforcer occurred within a common sequence of behavior. Zimmerman (1969), for instance, gave rats the opportunity to press one lever in order to obtain the presentation of a stimulus light on a fixed interval. Once that stimulus was presented, a response on a second lever would result in the delivery of food. The stimulus light, via its forward pairings with the food, should have accrued associative strength over the course of performance. Because, however, responding on the first component of the chain also activated the second manipulandum in the chain, it is difficult to assert that the animal was responding for the stimulus rather than the opportunity to respond on that second manipulandum. Chain schedules, therefore, typically require some further intervention in order to partition the sources of support for instrumental responding. In this case, Zimmerman took advantage of the fact that the pattern of responding on fixed interval and variable interval schedules differs to assess whether the light was controlling performance on the first lever as a secondary reward.
To do this, he put the rewarding impacts of the light and food into competition with each other on the first lever. In a test phase, the light was presented as a result of responding on the first lever on a variable interval schedule, whereas the food was presented on the fixed interval schedule that had previously delivered the light and the second lever was shifted to an extinction schedule. Zimmerman found that the pattern of responding on the first lever shifted from that typical of a fixed interval schedule to that typical of a variable interval schedule, a finding consistent with the development of conditioned reinforcing properties by the stimulus light.

Although commonly employed, the difficulty of ruling out alternative interpretations of the source of instrumental performance on chain schedules leaves something to be desired. Since the second response, as in Zimmerman’s study (1969), often becomes superfluous in the critical phase testing for the presence of SdR, it follows that it may not be necessary at all. Extinction studies of SdR reify this possibility, by utilizing a training phase where an instrumental action is paired with a stimulus that is immediately followed by the delivery of a reward. Because of the presence of the reward during training, the second, third, and higher components of the instrumental chain used to provide further conditioned stimuli and eventual primary reinforcement in chain studies of SdR are eliminated from the outset. A test phase is again required to detect the role of the SdR in the maintenance of that instrumental behavior. If the SdR plays no role in the maintenance of instrumental responding, then with or without its presence at an extinction test phase, animals should extinguish at the same rate. Instead, researchers usually find that animals extinguish much more slowly when the instrumental response leads to the delivery of the putative conditioned reinforcer than when it leads to no stimulus consequences (Bugelski, 1938).

Although these studies appear to confirm the basic effect, the most direct way to demonstrate and compare the secondary or sensory reward value of some event or other is to assess its ability to serve as the goal during the acquisition of a new action. If stimuli acquire the ability to reward instrumental actions in the course of pairing them with a primary reward, then it follows that one should be able to demonstrate the acquisition of instrumental actions that have as their sole outcome the delivery of a stimulus with a history of this pairing. This logic has been frequently employed in the detection of SdR, and procedures employing it have generally been referred to as acquisition of a new response, or simply, acquisition tests of SdR. Especially attractive is the absence of confounding effects of primary reward during training that could interfere with SdR interpretations of instrumental behaviors (Wike, 1966). Numerous experiments along these lines have been conducted by giving prior stimulus–outcome associations followed by training on a lever that delivers that stimulus. Work by Trevor Robbins and colleagues has demonstrated particularly clear acquisition of lever-press performance when that lever delivered a stimulus that was previously associated with food relative to an inactive lever that the rats could press but that had no scheduled consequences (Taylor and Robbins, 1984; Robbins et al., 1989).

We have conducted a similar experiment to those of Robbins using two different versions of their procedure, firstly to replicate their basic result but also to examine the effects of using a different control condition in which one lever delivered a stimulus that had previously been paired with food and the other lever delivered a familiar stimulus but that had not been paired with any rewarding consequence; a sensory reinforcement control (SeR). The results of this study are presented in Figure 3. As is clear from this figure, a good conditioned reinforcement effect was observed in both conditions: responding on the lever delivering the SdR was greater than on the inactive lever (left panel) and greater on the lever delivering the SdR than on the lever delivering the SeR. It is also clear, however, that the net size of the SdR effect is really much smaller than one might be led to believe from the difference between the active and inactive levers.

1.36.3.2 Do Secondary Rewards Reward, Reinstate, or Reinforce?

Describing events associated with primary reward as SdRs suggests that the responses that animals learn to gain access to SdRs are goal-directed. This is, however, a matter of dispute. It has often been argued in the past that, rather than developing reward value, the stimulus acquires the ability to drive instrumental responding in an S–R fashion, i.e., rather than acting as a goal in and of itself, it acts to reinforce the connection between situational cues and the response. That the conditioned reinforcing stimulus itself might not be the object of an instrumental action, but rather an elicitor of that action, is an explanation that has seen some theoretical and experimental exploration. Bugelski (1956), for instance, reinterpreted his
earlier extinction test-conditioned reinforcement data (Bugelski, 1938) using this framework and found that it provided a satisfactory account of the results. During acquisition training, the SdR not only follows the instrumental response as a consequence, but on all trials except the first bears a forward predictive relationship with later occurrences of that response. It is at least possible that during acquisition the instrumental action is reinforced solely by the primary reinforcer, whereas the SdR becomes associated with the response itself. In extinction, it is argued, the conditioned reinforcer then acts, following the first response, to delay extinction through this conditioned ability to evoke or reinstate subsequent instrumental responses.

Wyckoff (1959) attempted to produce a quantitative model of SdR effects emphasizing the eliciting function, or cue properties, of the conditioned reinforcer. This model was based on the results of an experiment reported by Wyckoff et al. (1958) in which rats were given conditioning trials where a buzzer was followed by the delivery of water. Following this training, experimental rats were given the opportunity to press a lever in order to secure the delivery of the buzzer without water, and control rats were placed on an omission schedule, where the buzzer was delivered if they refrained from pressing the lever. Performance between the two groups was not reliably different, which led Wyckoff et al. to conclude that the buzzer functioned primarily not to reward lever pressing, in which case the experimental group should have pressed significantly more than the control animals, but to elicit lever pressing. This result, however, has not been replicated, suggesting that some feature of the experimental design, or a simple lack of power, prevented Wyckoff et al. from observing cue-independent conditioned reinforcing effects. Indeed, Ward (1960) conducted a formally very similar experiment, substituting food reward for water and the random delivery of the cue in the control group for the omission schedule, and demonstrated a reliably greater level of responding in experimental animals than in control animals.

An important source of evidence against the response elicitation account of SdR effects comes from an experimental series performed by Crowder and his colleagues. They employed a yoked control procedure in several different paradigms to demonstrate the existence of secondary reward above and beyond the effects of stimulus-based response elicitation. In all experiments, the experimental animals performed an instrumental action that was followed by the delivery of SdR. At the same time as that delivery, the yoked controls received noncontingent presentation of that same stimulus. If the stimulus elicited or reinstated further responding, it should have done so equally in both groups. Instead, in the extinction test paradigm (Crowder et al., 1959a), the acquisition of a new response paradigm (Crowder et al., 1959b), and in reacquisition (Crowder et al., 1959c) and retention SdR paradigms (Crowder et al., 1959d), they found superior performance in the experimental subjects whose actions were correlated with the delivery of the SdR. Although these results are not completely immune to criticism derived from the analysis of systematic sources of error in the

Figure 3

An assessment of secondary reward conducted in a choice test on two levers. Rats were first given pairings between one stimulus (S1) and a rewarding outcome (O1), whereas another stimulus was presented unpaired (S2). Some rats were then allowed to press two levers. In one group, one lever delivered S1 and the other S2 (left panel) in the other group, one lever delivered S1 and the other nothing (Ø). It is clear that both tasks revealed a secondary rewarding effect of S1 on performance. However, the effect is somewhat exaggerated by the choice between S1 and Ø. When sensory reward is taken into consideration, as it is in the choice between S1 and S2, the net secondary reward effect is significantly smaller.
yoked group (Church, 1964), they indicate the relatively small degree of support that elicitation or reinstatement accounts provide for instrumental responding in SdR paradigms.

One published study has attempted to assess whether lever pressing for the SdR is goal-directed by devaluing the primary reward previously associated with the SdR (Parkinson et al., 2005). In this study rats were given pairings of a light stimulus paired with sugar after which the sugar was paired with illness. Although this reduced responding to the sucrose, it did not affect the ability of the light to serve as a secondary reward for lever pressing; lever pressing was acquired and maintained to a comparable degree whether the sucrose had been devalued or not. The authors concluded that, as the lever pressing appeared to be independent of the value of the primary reward, performance acquired through SdR should be considered habitual. But what was not confirmed in this study, however, was whether the devaluation of the sucrose was successful in modifying the reward value of the light. Indeed, as the SdR value depends on the association of the light with the emotional response elicited by the sucrose, rather than by the sucrose itself (see Figure 4), it seems unlikely that SdR could be undermined in this way. Rather, what this account predicts is that devaluation of the SdR could only be induced by counterconditioning, i.e., pairing the light previously paired with sucrose with a noxious consequence, such as foot shock. Would lever pressing still have been maintained after this treatment? To date no studies along these lines have been conducted, although there is plenty of evidence from studies of conditioned punishment to conclude that at least the sensory rewarding component of stimuli is abolished by this means of devaluation (Killcross et al., 1997).

Other studies from our laboratory have, however, confirmed that actions acquired for a secondary reward are essentially goal-directed. As discussed with respect to primary reward, one of the criteria for defining an action as goal-directed is that it is sensitive to the causal relationship between the action and reward. In this experiment, rats were first given pairings between two distinct visual cues with one cue paired with sucrose and the other with food pellets. After this training, the rats were trained to press two levers, each associated with a different visual cue. In these sessions, one of the visual cues was also presented noncontingently; as such the noncontingent cue was the same as that presented contingent on pressing one lever but different from that presented for pressing the other lever. As such the specific R-SdR contingency was maintained on one lever but was degraded on the other. The results of this study are presented in Figure 5. As is clear from this figure, the rats were sensitive to the specific lever press–SdR contingency, reducing performance on the action delivering the same SdR as that delivered noncontingently relative to the other action. This result is not consistent with either the reinforcing or reinstating functions of SdRs (Winterbauer, 2006).

An important aspect of the establishment of a secondary reward is its pairing with primary reward. The procedures that establish SdRs are, in fact, identical to those used to establish Pavlovian CRs to a stimulus. The possibility that stimuli require something more than Pavlovian conditioning to become conditioned reinforcers has been entertained; Skinner (1938) proposed, for example, in a thesis later considered in detail in the work of Keller and Schoenfeld (1950), that only stimuli that act to set the occasion for responding to other Pavlovian stimuli could serve as SdRs. Again, as Wike (1966) suggested, although occasion setters may make better SdRs, there seems to be no requirement that all conditioned reinforcers be occasion setters. Work in our laboratory has largely confirmed this view, in that we show perfectly reasonable SdR effects without special modifications to the Pavlovian conditioning phase (Winterbauer, 2006). But this raises an important issue: if SdR can be established using Pavlovian procedures, are the processes underlying reward and those underlying Pavlovian incentive motivation all one and the same? Or is this no more than a superficial, procedural similarity?
1.36.4 Reward and the Anticipation of Reward

The preceding sections have reviewed the considerable evidence suggesting that the influence of changes in reward value on goal-directed instrumental actions is an important determinant of action selection and of instrumental performance generally. Other factors can clearly influence performance, however. One of the most obvious, and perhaps best-documented, influences on action selection is that produced by stimuli ‘associated’ with reward. Advertising has a clear influence on action selection; if it did not the advertising industry would be a vacuous waste of time and of advertisers’ money. Of course, advertisers are hoping that the stimuli that they associate with a particular product will provide the basis for quite specific changes in choice performance and, of course, by and large they do. It is important to recognize, however, that, despite a superficial similarity in some of the procedures used to establish the reward value of particular events, notably SdRs, there is substantial evidence suggesting that the influence of cues associated with reward on goal-directed instrumental actions is not mediated by the reward system. In this section, we describe this evidence as it has emerged from analyses of the relationship between Pavlovian and instrumental conditioning, particularly those proposing that the motivational processes engaged by reward and by the anticipation of reward are the same or, at the very least, interact with one another.

1.36.4.1 Pavlovian-Instrumental Interactions

In fact, some of the earliest evidence that the representation of the instrumental outcome takes part in action selection was found by studying how Pavlovian and instrumental learning processes interact. For instance, Trapold (1970) trained rats on a biconditional discrimination in which, on any given trial, subjects were allowed to choose between two actions (left and right lever press). Trials were initiated by the presentation of one of two discriminative stimuli (tone and clicker), signaling which of the actions would be rewarded (e.g., S1 \rightarrow R1 and S2 \rightarrow R2). The novel feature of this experiment, however, was that these cues also signaled the identity of the outcome that could be earned on that trial. Whereas the control groups earned either food pellets (food control) or sucrose solution (sucrose control) on both actions (e.g., S1 \rightarrow R1 \rightarrow O1 and S2 \rightarrow R2 \rightarrow O2), the experimental group was rewarded with one outcome (O1; e.g., pellets) for performing one action and a different outcome (O2; e.g., sucrose) for performing the other action (e.g., S1 \rightarrow R1 \rightarrow O1 and S2 \rightarrow R2 \rightarrow O2). Consequently, the experimental group differed from the control groups in that, for the former, each discriminative stimulus signaled not only a different response but also a different outcome. Interestingly, Trapold (1970) found that the experimental group acquired more rapidly than either control group despite the fact that the S–R arrangements needed to solve the discrimination were the same across conditions.

This phenomenon, known as the differential outcomes effect, provides clear evidence that reward expectations can be used to guide action selection.
Moreover, the representation mediating this effect appears to consist of richly detailed information about the sensory properties of the reward. In Trapold’s (1970) study, the sucrose solution and grain-based pellets used to differentially reward the two actions were both nutritive outcomes and so should have held a similar incentive value for hungry rats. Since this motivational variable does not appear to have been used to discriminate between actions, rats probably relied instead on the sensory features (e.g., texture, odor, taste) of the anticipated outcome. There is even evidence that this effect can be obtained using outcomes that differ in one motivationally irrelevant sensory feature. Fedorchak and Bolles (1986) trained thirsty rats on a biconditional lever press discrimination task in which each correct response was rewarded with water. For two groups, the delivery of water was occasionally paired with a flashing light; whereas the light exclusively followed just one of the two S–R arrangements in the differential outcomes group, it followed both responses with an equal probability in the nondifferential control group. For a third group, the light was never paired with water. Once again, the group that received differential outcomes acquired more rapidly than the other two groups, demonstrating that the expectancy of a sensory event extraneous to outcome itself could be used to guide action selection.

How does differential outcomes training provide an advantage in discriminating between two actions? Clearly, it must have something to do with the Pavlovian contingencies embedded in the task (see Figure 6, top panel). It has long been argued that Pavlovian learning plays an important role in the control of instrumental performance (Rescorla and Solomon, 1967). Although we will discuss alternative accounts shortly, let us first consider the model Trapold and Overmier (1972) devised to explain the differential outcomes effect and similar findings (see Figure 6, middle panel). Their model was built within the general framework of traditional S–R theory (Hull, 1943), and so instrumental learning was assumed to involve the gradual recruitment of S–R associations through a conventional reinforcement process. However, Trapold and Overmier (1972) proposed that reward deliveries engage a second, Pavlovian learning process capable of supporting the acquisition of stimulus–reward associations. It was argued that through such learning, stimuli acquired the capacity to elicit a reward expectancy comprising the sensory features of that event. The final step in their argument was in allowing this reward expectancy to enter into S–R associations like any other sensorial event in the training environment, i.e., the expectation of reward was assumed to acquire discriminative control over performance. According to this analysis, the experimental group in Trapold’s (1970) study was provided with an additional source of stimulus support for action selection; the correct choice was signaled by both an auditory cue and an expectation of the reward that could be earned on that trial.

The differential outcomes effect provides strong evidence that the Pavlovian learning can influence instrumental performance through a highly specific representation of the mediating outcome. Further evidence for this claim comes from studies of the so-called Pavlovian-instrumental transfer effect. For instance, Kruse et al. (1983) first trained rats using a biconditional procedure quite similar to that used in differential outcomes studies (e.g., Trapold, 1970;
Fedorchak and Bolles, 1986), such that each stimulus (clicker and tone) signaled both the response (left or right lever) that would be rewarded and the identity of its outcome (pellets or sucrose solution). During a separate Pavlovian training phase, the group of interest to our current discussion received pairings between a stimulus (pause in the background white noise) and one of the two outcomes (either pellets or sucrose). In a subsequent test phase, Kruse et al. (1983) found that presentations of this stimulus facilitated instrumental performance in an outcome-dependent manner; rats preferentially increased their performance of the action that had shared a common outcome with that cue, relative to the other action. Importantly, this test was conducted in extinction, indicating that this effect relied entirely on information acquired during earlier training phases.

Following Kruse et al. (1983), there have been numerous demonstrations of outcome-specific transfer (Colwill and Rescorla, 1988), even using actions that had been acquired through free operant training (Colwill and Motzkin, 1994; Delamater, 1995; Holland, 2004). The latter finding is important because it reveals that Pavlovian learning can influence action selection even under conditions in which anticipating reward provides no obvious advantage in obtaining reward. According to Trapold and Overmier (1972), the transfer effect emerges because the Pavlovian outcome expectancy selectively retrieves the response it signaled during training through the activation of an outcome–response association (see Figure 6, top panel). This account applies equally well to the free operant situation. Note that in any instrumental conditioning study there exists an embedded Pavlovian relationship between contextual cues and the reward delivery. In this case, cues that best predict reward should come to elicit an expectancy of reward capable of entering into association with the response.

Of course, the two-process account of Trapold and Overmier (1972) does not provide the only explanation for the influence of Pavlovian reward expectancies over instrumental performance. For instance, several two-process theories have been proposed that assume instrumental learning involves encoding some approximation of the action–outcome contingency arranged by the experimenter (Bolles, 1972; Asratyan, 1974). According to this view, Pavlovian outcome expectancies guide action selection by retrieving the action that had actually earned that outcome during training (see Figure 6, bottom panel).

### 1.36.4.2 The Two-Process Account of Reward Value

How is Pavlovian-instrumental transfer relevant to our interpretation of instrumental performance as an instance of goal-directed action? Recall that in order to be considered goal-directed, a behavior must be performed because of its expected consequences; performance should depend on the subject’s capacity to (1) anticipate the outcome of the action (i.e., action–outcome learning) and (2) evaluate the incentive properties of that outcome (i.e., incentive learning). Two-process theories, however, tend to attribute incentive effects, such as the sensitivity of instrumental performance to outcome devaluation, to the Pavlovian process (Rescorla and Solomon, 1967). These accounts typically assume that Pavlovian learning provides the motivational support for instrumental performance. Even Trapold and Overmier (1972), who took an expressly associative approach, entertained the possibility that incentive manipulations have their effect by disrupting the capacity of the Pavlovian outcome expectancy to mediate response selection (e.g., through generalization decrement). Others have taken a more explicitly motivational position. Bolles (1972), for instance, proposed that Pavlovian and instrumental processes interact based on their shared outcome expectancies, but that this interaction is gated by the incentive value of mediating outcome. The two-process approach, therefore, provides a compelling explanation for the influence of reward value over performance. According to this account, instrumental responding is depressed following outcome devaluation, not because of a reduction in the reward value of the outcome and knowledge of the underlying response–outcome contingency, but because this treatment diminishes the Pavlovian support for performance.

The claim that Pavlovian learning plays a part in action selection is beyond doubt. The critical question, however, is whether these processes are responsible for the influence of reward value over performance. If so, it would be necessary to abandon the goal-directed interpretation of instrumental performance altogether. Note that since the two-process account uses the Pavlovian–instrumental interaction responsible for transfer to explain the sensitivity of performance to outcome devaluation, it predicts that these two apparently distinct forms of action selection should share a common associative structure. One way to evaluate this prediction, therefore, is to assess whether the associations guiding
transfer and outcome devaluation are acquired at roughly the same rate. For instance, it has been repeatedly shown that, while sensitivity to outcome devaluation emerges with rather limited training (Holland, 2004; Yin et al., 2005), depending on training parameters used (e.g., number of action–outcome contingencies), this effect is either maintained (Colwill and Rescorla, 1985a) or attenuated (Adams, 1981; Holland, 2004) with training that is more extensive. Alternatively, recent evidence suggests that Pavlovian-instrumental transfer increases in magnitude with more extensive instrumental training (Holland, 2004).

It should also be possible to evaluate the two-process account by analyzing the content of the associations that mediate transfer and outcome devaluation. However, it is important to remember that individual two-process theories do not agree on what that content should be. Trapold and Overmier (1972), for instance, argued that the response becomes associated with an expectancy of reward generated by prevailing stimuli, resulting in an outcome–response association. As we have mentioned, others (e.g., Asratyan, 1974; Bolles, 1972) have proposed that the response becomes associated with the outcome it actually produces during training, in the form of a response–outcome association. Two-process theories, therefore, can be distinguished by determining whether the association responsible for action selection reflects the actual response–outcome contingency, or whether it is, instead, the product of the incidental stimulus–outcome contingency present during training. However, investigating the relative contribution of these two contingencies to instrumental learning is no trivial task. In any typical instrumental conditioning study, the outcome earned by the response is also predicted by the prevailing situational cues (i.e., the anticipated and earned outcomes are the same). Thus, one approach to the problem is to create a training situation in which this in not the case.

Several studies have used this basic strategy to assess the associative structure underlying transfer and outcome devaluation. For instance, Colwill (1994) reported evidence of outcome selective transfer with responses that had been concurrently trained on distinct action–outcome contingencies. Similarly, Colwill and Rescorla (1985b) reported that rats display an outcome-specific devaluation effect after concurrent training of this kind. Since rats given concurrent training are allowed to alternate freely between responses, the context should be associated equally with both outcomes, thereby preventing the development of specific outcome–response associations. The specificity of transfer and outcome devaluation despite this treatment, therefore, seems to suggest that both effects can be supported by response–outcome learning.

Rescorla and Colwill (1989) and Rescorla (1992) have attempted more directly to compare the relative contribution of outcome–response and response–outcome associations to these effects. For instance, Rescorla and Colwill (1989) investigated this issue by first pretraining rats on a common nose-poke response with four distinct stimuli; two stimuli (S1 and S3) signaled a pellet reward and two others (S2 and S4) signaled a sucrose solution. Next, they were given discrimination training on two responses (R1 and R2), such that one response, say R1, earned pellets and the other response, R2, earned sucrose. However, each response was also signaled by a stimulus that had previously been paired with the alternative outcome (i.e., S2 → R1 → O1 and S1 → R2 → O2). According to Trapold and Overmier’s (1972) two-process account, this should have resulted in the formation of, for example, a sucrose–R1 association, even though R1 had actually been followed by pellets. During the transfer test, rats were allowed to perform each response in extinction while S3 and S4 were occasionally presented. In contrast to the predictions of the outcome–response view, it was found that stimulus presentations selectively facilitated performance based on the identity of the outcome that ‘followed’ a response during training (e.g., S3 increased R1 relative to R2). Furthermore, in a separate experiment, Rescorla and Colwill (1989) used the same strategy to investigate the structure underlying outcome devaluation performance. They found that, as with transfer, the sensitivity of instrumental performance to reward value was dominated by response–outcome learning; performance was suppressed by devaluing the outcome that the action had actually earned during training, not the outcome that was signaled by the discriminative stimulus (e.g., devaluing O1 decreased R1 relative to R2).

There is, however, reason to question whether these experiments provide a fair test of the outcome–response account. This basic approach, of course, depends entirely on the experimenter’s capacity to create a situation in which the expectation of reward differs from the reward that is obtained by responding. In Rescorla and Colwill’s (1989) study, for instance, each discriminative stimulus was pre-trained so that it would signal a different outcome.
from the one that would be earned on that trial. Since this phase of the experiment was conducted over 4 days, however, it is possible that rats were able to learn the new stimulus–outcome relationships (e.g., \(S_1 \rightarrow O_2\)), nullifying the effects of pretraining. Rescorla (1992) addressed this issue in an experiment otherwise quite similar to the first (Rescorla and Colwill, 1989), except that, during the discrimination phase, each stimulus continued to be paired with the outcome that it predicted during initial pretraining, while at the same time signaling that responding could earn the opposite outcome. These additional Pavlovian trials were added to encourage the persistence of the initial stimulus–outcome learning, thereby providing greater opportunity for any potential outcome–response associations to form during the instrumental discrimination training. Using this new procedure, Rescorla (1992) once again found no evidence that outcome–response associations play a part in outcome devaluation performance. However, the results of transfer testing were less straightforward. He observed that stimulus presentations tended to increase the performance of both responses, although this effect was larger for the response that had ‘earned’ the outcome signaled by the transfer stimulus than it was for the response that had been trained in ‘anticipation’ of that outcome. Thus, while these findings suggest that both outcome devaluation and transfer are dominated by response–outcome learning, they also indicate that outcome–response associations may play some, albeit limited, role in the latter.

This conclusion does not help the two-process account of reward value. According to this account, the processes underlying transfer and outcome devaluation should be identical. Perhaps more importantly, however, these studies illustrate the difficulty in attempting to dissociate the contributions of Pavlovian and instrumental learning to performance. Indeed, even in these studies it is possible that the subjects were able to confound the experimenter’s intentions and acquire appropriate stimulus–outcome associations during instrumental training based on the relationship between the features of the individual response manipulanda and the outcome earned by those responses. For instance, rats trained to press a lever for pellets and pull a chain for sucrose solution may come to associate the lever itself with pellets and the chain with sucrose. Such learning would ensure that the rat anticipated the reward that they would actually obtain for performing the response, even in the presence of a context that signaled both rewards (e.g., Colwill and Rescorla, 1985b; Colwill, 1994) or a Pavlovian cue that signaled a different reward (e.g., Rescorla and Colwill, 1989; Rescorla, 1992).

This problem can be avoided, however, by training distinct action–outcome contingencies on a common response manipulandum. For instance, Dickinson et al. (1996) trained rats to push a vertically positioned pole to the left and right for different outcomes; for half the rats, left pushes earned food pellets and right pushes earned a maltodextrin solution, whereas the other half was trained with the opposite arrangement. Rats were then sated on one of the two outcomes in order to selectively reduce its reward value. Immediately after this treatment, they were given an extinction test in which the pole was available and could be pushed freely in either direction without consequence. Dickinson et al. (1996) found that, despite having both actions trained on a common manipulandum, the rats were able to use response–outcome training relationships to guide their action selection according to outcome value; rats were less likely to push the pole in the direction that had earned the now devalued outcome, relative to the other direction. This finding is incompatible with the two-process account, which predicts that outcome-selective devaluation should never emerge in the absence of differential stimulus–outcome contingencies. Instead, it provides strong support for the view that instrumental performance is goal-directed and that its sensitivity to reward value depends on response–outcome learning.

One final method for evaluation of the two-process account of reward value involves assessing the interaction between transfer and outcome devaluation. If these phenomena rely on the same underlying structure, then the capacity of a Pavlovian cue to facilitate performance should depend on the value of the mediating outcome representation. Colwill and Rescorla (1990) directly investigated the role of incentive value in outcome selective transfer. Rats were initially given biconditional discrimination training using differential outcomes, such that one stimulus (\(S_1\)) signaled that pellets could be earned on one response (\(R_1\)) and the other stimulus (\(S_2\)) signaled that sucrose could be earned on a different response (\(R_2\)). Subsequently, they were given free operant training on two new responses (\(R_3\) and \(R_4\)), such that each earned a unique outcome (either pellets or sucrose). One outcome was then devalued through
conditioned taste aversion and then a transfer test was conducted in extinction, with both R3 and R4 available. Although rats were, in general, less likely to perform the response that had earned the devalued outcome than the other response, both responses were selectively facilitated by presentations of the stimulus with which they shared a common outcome. Moreover, the magnitude of this transfer effect, measured in the difference from baseline performance, was comparable across responses. This basic finding, that devaluing an outcome fails to diminish its capacity to mediate Pavlovian-instrumental transfer, has since been replicated in a number of studies (Rescorla, 1994; Holland, 2004).

Altogether, there appears to be scant support for the two-process account of reward value. The associative processes supporting outcome devaluation and transfer appear to be acquired at different rates and encode somewhat different content. Furthermore, instrumental responses remain sensitive to outcome devaluation under conditions that cannot support differential stimulus–outcome learning. Finally, the Pavlovian–instrumental interaction responsible for transfer does not appear to depend on the reward value of the retrieved outcome. Instead, these findings strengthen the goal-directed view of instrumental action and, while demonstrating that reward anticipation influences action selection, it is also clear that this effect is not mediated by the reward system.

### 1.36.5 Summary and Conclusions

We have argued that the reward system is a specialization that developed in the service of goal-directed action allowing animals to encode the relative values of specific environmental events. These values provide the basis for choice, allowing animals to decide on a course of action based not only on knowledge or information as to the consequences of an action but on the basis of the value of those consequences.

Encoding the reward value of a particular event involves the formation of an association between the specific sensory representation of that event and an emotional response. In the case of primary rewards, the emotional response is directly determined by the activity of specific motivational and affective processes engaged during consummatory contact with the outcome. Thus, by virtue of their biologically active properties (e.g., nutrient, fluidic, pheromonal), rewarding events (food, fluid, sex objects, and so on) are readily able to activate these underlying systems that modify emotional responses as one of the consequences of that activation. Basing the evaluation of primary rewards on emotional responses is adaptive if those responses are determined by the operation of these basic motivational and affective systems, which is essential if the animal's choice between alternative courses of action is to remain, by and large, adaptive too. In the case of secondary rewards, the emotional response is, of course, determined by the primary reward with which it is paired. By basing the transfer of value from primary to secondary rewards on an emotional response, the selection of actions, even when they are directed toward achieving apparently quite arbitrary goals, can be understood as being constrained by primary motivational processes through their influence on emotional responses.

Finally, we addressed the distinction between the role the reward system plays in assigning reward value and the processes controlling the anticipation of reward. These are quite distinct aspects of behavioral control; although cues that signal forthcoming rewards can provide information that can be used by the goal-directed system, they do not depend, ultimately, on the reward system to play that role. As such, the influence of reward-related cues on action selection does not replace or explain away the functions of the reward system in this regard. Rather, the distinct processes mediating the effects of reward and of the anticipation of reward provides the basis for understanding the role that cognitive processes generally play in goal-directed action. Because it constrains the event relations to which an animal is exposed, there has been a long tradition of using Pavlovian conditioning to model the cognitive control of behavior. The fact that, ultimately, this system is concerned with the production of reflexive responses would, however, appear to render this approach perhaps a little too abstract. It makes more sense to study the role of cognition in a behavioral system within which information can act to influence performance. Based on the evidence reviewed here that animals are able to exert control over their instrumental actions, choose between actions based on the relative value of their consequences, and use predictive information to influence action selection, we suggest that instrumental conditioning provides the more precise model of this capacity.
Acknowledgments

The preparation of this chapter was supported by the National Institute of Mental Health, grant #56446.

References


