Published in final edited form as:
Drug Discov Today Dis Models. 2008 ; 5(4): 235–245. doi:10.1016/j.ddmod.2009.07.004.

On habits and addiction: An associative analysis of compulsive drug seeking

Sean B. Ostlund[1] and **Bernard W. Balleine**[1,2]

[1] Departments of Psychology, Psychiatry and Biobehavioral Sciences and the Brain Research Institute, University of California, Los Angeles

[2] Brain & Mind Research Institute, University of Sydney

Abstract

The processes that underlie the pathological pursuit of drugs in addiction and that support the transition from casual drug taking to their compulsive pursuit have recently been proposed to reflect the interaction of two action control processes that mediate the goal-directed and habitual control of actions for natural rewards. Here we describe the evidence for these learning processes, their associate structure and the motivational mechanisms through which their operation is translated into performance. Finally, we describe the potential changes in the interaction between habitual and goal-directed processes induced by drug addiction that subserve compulsive drug pursuit; i.e. the increase in habit learning and reduction in the regulation of habits induced by changes in the circuitry that mediates goal-directed action.

Keywords

Instrumental conditioning; Pavlovian conditioning; goal-directed action; habit; dorsal striatum

Drug addicts pursue and use drugs despite their often extensive, first-hand experience with the numerous adverse effects of this behavior. Understanding how drugs of abuse can lead to the development of compulsive drug seeking and taking is therefore one of major goals of addiction research. An increasingly popular approach to this problem views addiction as a dysfunction of learning [1,2]. From this perspective, drugs of abuse not only engage the neural systems that support normal learning and memory, they are particularly effective in doing so, leading to aberrations in the strength or content of learning.

Rather than attempting to provide a general overview of this literature, which can be found elsewhere [3], the current paper will focus on one hypothesis in particular – that drug addiction results from a dysregulation in the formation and/or execution of habits [4]. Of course, this basic notion has a long history in popular culture. Cigarette smoking, for example, is often described as a ‘bad habit’, and so quitting smoking should merely be a matter of ‘breaking’ that habit. However, as any current, former, or relapsing smoker will tell you, this view is an over-simplification that trivializes the problem of addiction. However, as we shall see, the compulsive nature of drug seeking/taking is nicely captured by the concept of habit learning and empirical evidence for this hypothesis has rapidly accumulated in recent years.

Correspondence. Bernard W. Balleine, Department of Psychology, UCLA, Box 951563, Los Angeles, CA 90095-1563, Tel: 310 825-7560 (office), 310 825-2998 (lab), Fax: 310 206-5895, balleine@psych.ucla.edu.

The aim of the current paper is to develop a theoretical context for thinking about the role of habit learning in the addiction process. Along the way, we hope to identify problems inherent in modeling compulsive drug seeking/taking in animals and refine the predictions of this general hypothesis in light of what we have learned about habit learning from instrumental conditioning studies using natural rewards.

What is habit learning?

In the field of learning theory, the term habit tends to be reserved for the product of stimulus-response learning [5,6]. Consider an instrumental conditioning experiment in which a hungry rat is rewarded with food pellets for pressing a lever. In general, theories of stimulus-response learning assume that these food pellet deliveries strengthen, or *reinforce*, an association between prevailing stimuli (e.g., the experimental context or other more punctate stimuli such as the sight of the lever) and the lever press response. Consequently, when returned to the training situation, the rat should automatically perform the lever press response; no actual planning is necessary, of course, because the mediating associations have already been established. Conversely, it is assumed that aversive events, like an unpleasant food or an electric footshock, can serve to punish responding resulting in the weakening of stimulus-response associations.

A critical tenant of stimulus-response theory is the contention that the reward is not encoded as a component of the associative structure supporting instrumental performance; it serves only as a catalyst to strengthen the S-R association and is not encoded as a consequence (or goal) of the action [7]. However, this tenant leads to an important prediction: although the motivational value of a response-contingent event (i.e., a reinforcer or punisher) should influence the initial acquisition of stimulus-response associations, it should not impact the expression of habitual performance. That is, a habitual response should be elicited by its associated stimuli and so performed even if, in the interim, the consequences of that response have become noxious or aversive. Habits can, therefore, be contrasted with actions that are performed in a more flexible, goal-directed manner. Unlike habits, which are performed in spite of their consequences, goal-directed actions are performed because they are expected to produce some desirable outcome (or are withheld because they are expected to produce some unpleasant outcome). Given this perspective, it is easy to see the appeal of the habit learning account of compulsive drug use; if stimulus-response learning underlies the execution of drug-related behaviors, then the performance of these behaviors should be insensitive to their consequences relative to the addict's other goal-directed behavior.

At this point, however, it is worth considering the validity of this distinction between habitual and goal-directed actions. Of course, both accounts adequately explain why a hungry rat will press a lever that has been rewarded with food pellets; either it presses because this response has been elicited by some environmental stimulus (i.e. via a stimulus-response association) or because it wants food pellets and has learned that they can be obtained by pressing the lever (i.e., via an action-outcome association). To establish whether any specific action is habitual or goal-directed, therefore, one needs a way of discriminating between habitual and goal-directed control of performance.

Tests of goal-directed action selection

Some time ago, on the basis of analyses of goal-directed behavior derived from human action theory (e.g. [8]), Dickinson and Balleine [9] argued that, in order to be classified as goal-directed, the performance of an action must be shown to depend on two things: (i) the causal relationship between that action and its outcome (i.e., its action-outcome contingency) and (ii) the current motivational value of that outcome. It turns out that under certain training conditions instrumental lever pressing in rodents meets both of these conditions.

For example, it has been well documented that changes in action-outcome contingency can alter the rate at which rats perform an instrumental response [10,11]. The term action-outcome contingency actually refers to the integration of two different conditional probabilities: the probability that an outcome will be delivered if the action is performed (i.e., p(O/R)) and the probability that the outcome will be delivered non-contingently, in the absence of the action (i.e., p(O/~R)). Typically, the latter probability is simply subtracted from the former, resulting in a value reflecting the net strength of association between the action and outcome (i.e. ranging between −1 and +1) [12]. When an appetitive outcome is being delivered, the performance of a goal-directed action should be positively related to the p(O/R) and negatively related to the p(O/~R); i.e., it should reflect the degree to which the response is necessary for obtaining the reward. There are numerous examples of how changes in p(O/R) can impact instrumental performance. Perhaps the clearest and best characterized example of this effect is the extinction of performance that occurs when reinforcement is suspended (i.e., when p(O/R) is shifted from some positive number to zero). This finding, however, also illustrates how difficult it is to distinguish between habitual and goal-directed performance because, naturally enough, stimulus-response theories can also explain the extinction effect; performing a response in the absence of reinforcement is assumed to lead to a gradual weakening of stimulus-response associations. Therefore, to unambiguously assay action-outcome contingency learning one must assess the sensitivity of a response to the effect of non-contingent reward deliveries; i.e. a situation in which p(O/R) is maintained as p(O/~R) is increased – a manipulation referred to as action-outcome contingency degradation – because, in this situation, S-R theory does not anticipate any effect of p(O/~R). An illustration of this procedure is provided in Figure 1.

In one of now many demonstrations of this effect, Balleine and Dickinson [13] trained rats to perform two different responses (R1 and R2; i.e. lever pressing and chain pulling) for distinctive food outcomes (O1 and O2; i.e. grain pellets and polycose solution) such that R1→O1 and R2→O2. By the end of training both responses were rewarded on a random ratio (RR) 20 schedule of reinforcement meaning that each response performed earned its corresponding outcome with a probability of 0.05. The sensitivity of instrumental performance to non-contingent reward delivery was then assessed by arranging that the probability of earning that outcome by responding was the same as the probability of earning that outcome in the absence of responding; that the outcome was equally probable whether the response was performed or not (i.e., p(O1/R1) = p(O1/~R1) = 0.05). This was, however, only true for the R1-O1 relationship; performing R2 was still the only way for subjects to obtain O2. As predicted by the goal-directed account of instrumental performance, Balleine and Dickinson [13] reported that the rats were able to suppress their performance of the action whose underlying action-outcome contingency had been degraded (R1) while continuing to perform the other response (R2), whose contingency had not been degraded.

The outcome-selectivity of this effect is noteworthy. First, it provides a demonstration that the rats had indeed encoded quite specific representations of the two action-outcome relationships with which they were trained. Without these associations, the rats would have had no basis for selectively suppressing their performance of the action that earned the outcome that had also been delivered non-contingently. This selectivity also provides an important control against alternative interpretations of this effect. For instance, if the rats had been trained on a single response-outcome relationship before receiving non-contingent presentations of that outcome, one might worry that any corresponding decrease in the instrumental response was merely the product of response competition between that action and behavior related to the collection of the free reward deliveries (e.g., approaching and entering the food magazine). However, this response competition interpretation cannot explain the outcome-selective effect because any competing response should interfere equally with the performance of both instrumental actions.

The dependence of instrumental performance on the action-outcome contingency can also be assessed using an omission test, in which the performance of an action actually prevents the subject from obtaining rewards that would have otherwise been delivered [14]. Omission training is described, therefore, as generating a negative action-outcome contingency, because the probability of earning reward by responding is actually less than earning reward by not responding (i.e. $p(O/A) < p(O/\sim A)$). In this situation, of course, it makes sense to inhibit one's performance in order to maximize the delivery of reward. Consistent with this goal-directed analysis of instrumental performance, it has been well established that rats confronted with an omission contingency of this kind show lower response rates than rats given an appropriate yoked control treatment in which they receive the same number and distribution of free rewards as the omission group but have no control over these reward deliveries (i.e., they experience no contingency between their instrumental performance and reward) [14].

Although the instrumental performance of rats appears to be guided by action-outcome learning, at least in the situations described above, this does not, by itself, explain why an animal would actually perform an instrumental action. For instance, if a rat had learned that pressing a lever results in the delivery of a painful foot-shock but nevertheless goes ahead and presses the lever, we would rightly doubt that the performance of this response was goal-directed. To be regarded as goal-directed therefore, an action must not only respect the causal relationship between action and outcome, it must also respect the value of the outcome. Simply stated, if the anticipated outcome is desirable then a goal-directed action should be more probable; if the anticipated outcome is aversive then performance should be less probable.

Indeed, as noted above, it is the dependence of performance on the *current* value of the outcome that allows goal-directed actions to be distinguished from habitual responses [9,15,16]. Although the value of a reinforcer is assumed to determine its ability to strengthen S-R associations, the encoded relationship between the action and outcome plays no direct role in the expression of habitual performance and, hence, there is no mechanism for allowing changes in the value of the outcome to influence performance in the absence of performing the response and earning that newly valued outcome. Thus, in contrast to goal-directed instrumental performance, changing the value of an outcome between training and testing should have no effect on habitual instrumental performance. In fact there is an abundance of evidence that post-training changes in outcome value can affect instrumental performance (cf. [9,15,16] for reviews). This evidence has come from outcome devaluation procedure illustrated in Figure 1. For example, in the Balleine and Dickinson [13] study described above, rats were initially administered an outcome devaluation test before undergoing contingency degradation training. Again, the rats had been trained on two separate action-outcome relationships (R1-O1 and R2-O2). To assess the sensitivity of their performance to outcome devaluation, each rat was then given one hour of unlimited access to one of the two training outcomes immediately before a test session conducted in extinction; i.e. in the absence of either outcome. At test, the rats were found to withhold their performance of the response that, in training, had earned the now devalued outcome, while continuing to perform the response that had earned the other, non-devalued outcome.

This method of outcome devaluation – termed sensory-specific satiety – is a particularly effective treatment for inducing a short-term reduction in the incentive value of food outcomes [17]. Long-term changes in outcome value can, however, also be accomplished by conditioning a taste aversion to an outcome. This technique produces a similar suppressive effect on instrumental performance and is commonly used in such experiments [18]. These procedures are illustrated in Figure 1.

Pavlovian-instrumental interactions

As with the contingency degradation effect, the outcome-selectivity of outcome devaluation allows one to rule out alternative interpretations of the sensitivity of instrumental performance to the devaluation manipulation. In order to understand why this is the case, it is important to consider another role that contextual cues could play in instrumental conditioning (i.e., apart from their putative role as discriminative, or eliciting, stimuli in habit learning). In a typical instrumental conditioning experiment, the outcome is paired both with the action on which its delivery is contingent and with the training context. Thus, subjects have the opportunity to learn a Pavlovian context-outcome relationship. Over the years, many theorists have argued that such Pavlovian learning, although often incidental to the task arranged by the experimenter, plays a central role in the control of instrumental performance [19–22]. Many of these so-called *two-process theories* posit that Pavlovian learning provides the motivational support for habitual (i.e., stimulus-response mediated) instrumental responding (e.g. [21]). When the outcome is valuable, as in initial training, the context should provide robust motivational support. Since this support is assumed to depend on the current value of the outcome, devaluing the outcome should reduce that motivational support and result in an immediate suppression of instrumental performance. According to the two-process account, therefore, responding should be withheld, not because the rats are able to evaluate the consequences of their behavior, as is assumed by the action-outcome account, but because the context no longer predicts a valuable outcome. Now, if Balleine and Dickinson [13] had trained their rats on a single action-outcome association, we would not be able to tell whether any resulting decrease in performance was the product of an action-outcome mediated evaluation process or a lack of Pavlovian (context-outcome) support for performance. However, in their experiment, the two action-outcome relationships were trained in a common experimental context. As a result, the context at test should have been equally associated with both the devalued outcome and the non-devalued outcome. Hence, the two-process account predicts a nonspecific decrease in the performance of both actions, regardless of the current value of their actual outcomes. As such, the outcome-selectivity of the devaluation effect observed in this study provides unambiguous evidence of action-outcome encoding.

Although these and other experiments provide consistent evidence against two process theories of instrumental performance (e.g. [23,24]; cf. [19] for review), reward-paired cues are known to influence instrumental performance in number of other important ways. For instance, rats tend to persist in performing a (previously rewarded) response in extinction if doing so results in the delivery of CS that had been independently trained to signal reward [25]. This phenomenon, termed conditioned or secondary reinforcement, is also demonstrated by the finding that response-contingent CS deliveries can support the acquisition of an entirely new (i.e., untrained) instrumental response [25,26].

However, Pavlovian cues do not have to be delivered contingent on responding in order to influence instrumental performance. The Pavlovian-instrumental transfer effect, for example, refers to the finding that rats perform instrumental responses more vigorously in the presence of a non-contingently delivered CS than in the absence of such cues [27]. Interestingly, recent reports of neural and behavioral dissociations indicate that there are two fundamentally different forms of transfer: an outcome-specific form and a general form [28,29]. Outcome-specific transfer, as the name implies, is characterized by the tendency of a CS to selectively increase the performance of a response trained with the outcome predicted by the CS, relative to a second response trained with a different outcome [30]. In contrast, in general transfer, the CS tends to facilitate instrumental performance independently of the sensory features of the training outcomes used. These procedures are illustrated in Figure 2.

One important difference between these phenomena lies in their sensitivity to shifts in motivational state; whereas specific transfer is unaffected by (for example) a shift from hunger to satiety, the general transfer effect is completely abolished by this treatment [31]. This latter effect should, however, be interpreted with some caution. Although being shifted to a state of general satiety – which was accomplished, in these experiments, by allowing rats to feed freely on their home chow – may be viewed as a manipulation of outcome value, it is more likely an effect on general activation and, as such, this finding should not be taken to indicate a role for Pavlovian learning in outcome devaluation performance. Indeed, neither the outcome-specific nor the general form transfer is affected by treatments that specifically devalue the mediating outcome (but not other outcomes); e.g., the tendency of a CS to potentiate instrumental performance is not diminished by devaluating the outcome that it predicts through conditioned taste aversion [32,33]. This finding is made all the more striking by demonstrations that CS-evoked anticipatory behavior (e.g., magazine approach) is influenced by changes in outcome value [34], confirming that, although the CR depends on the value of the US predicted by the CS, the CS exerts control over instrumental performance through an independent motivational process.

Habit formation

As discussed above, the sensitivity of instrumental performance to manipulations of outcome value and action-outcome contingency degradation indicates that rats can apply a goal-directed strategy to control the performance of instrumental actions. Although this might be taken to imply that it is the R-O and not the S-R learning process that dominates instrumental learning, this is not always the case. Certain types of training are known to generate performance that is relatively insensitive to these manipulations, suggesting that such training encourages stimulus-response, or habit, learning. For example, habit formation can be induced by reinforcing an instrumental response on a variable interval (VI) schedule such that rewards are made available only after a specified period of time has passed since the last reward was earned [35]. Habitual performance can also be established by overtraining rats on a particularly response [7,36]. Adams and Dickinson [7], for example, found that the performance of rats trained to lever press on a continuous reinforcement schedule was sensitive to outcome devaluation if they were allowed to earn 100 outcomes, but became insensitive to this treatment after they were allowed to earn 500 outcomes. Similarly, it has been shown that giving rats extensive training on a response can render the performance of that response less sensitive to omission training [37], suggesting that overtraining fundamentally changes the associative structure underlying the control of instrumental performance, rather than, say, altering reward processing.

Interestingly, it has been reported that as responding is rendered less sensitive to outcome devaluation through overtraining it becomes more sensitive to the excitatory influence of Pavlovian cues; i.e., the tendency of CS to potentiate responding is greater for overtrained than for undertrained responses [32]. This finding reveals another way in which habit learning may result in compulsive behavior; not only is habitual performance, by definition, insensitive to outcome value and rigidly-dependent on environmental (discriminative) cues, it is also more strongly modulated by Pavlovian cues, which, as discussed above, appear to affect performance regardless of the current incentive value of the predicted outcome. From this perspective, the Pavlovian-instrumental transfer effect has much in common with compulsive drug-craving, thought to be a central factor controlling relapse in drug seeking.

As we have seen, the training procedures used to establish instrumental conditioning can be critical in determining whether habit (stimulus-response) or goal-directed (action-outcome) learning will dominate performance at test. Furthermore, there is now a considerable body of evidence that these learning processes are mediated by separate neural systems. For example,

pre-training lesions of the prelimbic region of the prefrontal cortex (PL) [13,38], the posterior dorsomedial striatum (pDMS) [39,40], the mediodorsal thalamus (MD) [41]or the basolateral complex of the amygdala (BLA) [29,42] have all been shown to abolish the impact of outcome devaluation and action-outcome contingency degradation on instrumental performance, indicating that each of these structures plays a critical role in goal-directed instrumental conditioning. In contrast, the dorsolateral striatum (DLS) [43,44] and infralimbic cortex (IL) [45] have been implicated in habit formation; disrupting the normal functioning of these structures increases sensitivity to both outcome devaluation and contingency degradation treatments in rats given training (i.e., overtraining or VI-training) sufficient to induce habits.

All of the outcome devaluation findings that we have described so far were conducted in extinction to assess performance without giving subjects any external feedback about the current value of the reward(s). Testing in extinction is critical when attempting to distinguish between goal-directed and habitual performance because only the former should show sensitivity to outcome devaluation in the absence of such feedback. However, even theories of stimulus-response learning predict that a response should be suppressed if it actually *earns* a devalued outcome. At the very least, a devalued outcome should lose its ability to reinforce responding and, indeed, if the aversive properties of the devalued outcome come to outweigh its appetitive properties, its delivery should be expected to reduce any previously established S-R association.

However, given the heterogeneous processes available to control instrumental performance, it is not unreasonable to suppose that, over and above, or perhaps in addition, to the effects of punishment on S-R association, control of performance might be expected to revert to a goal-directed strategy when habitual actions need to be suppressed in the face of negative feedback. In fact, in contrast to the relatively slow, trial-by-trial changes in performance that should be predicted on the S-R account, this reversion to a goal-directed strategy should be anticipated to result in a much faster change in performance; and indeed, what evidence we have suggests that it is [60].

Figure 3(a) displays the results of an experiment that assesses this issue in which rats were initially given training to press a lever for a sucrose solution on which they were either undertrained (UT) for 120 reinforced actions or overtrained (OT) for 480 reinforced actions. The results of the extinction test clearly show that performance generated by overtraining was impervious to outcome devaluation. As can be seen in the left panel of Figure 3(a), the undertrained group were sensitive to devaluation (DEV), here induced by taste aversion learning, and Group UT-DEV responded markedly less on the lever in the extinction test compared to Group UT-NON. In contrast, the overtrained groups did not differ on test; Group OT-DEV responded just as much on the lever as Group OT-NON. In this experiment, however, the subjects were given a second test in which responding actually delivered the now noxious sucrose outcome. As can be seen in the right panel of Figure 3(a), this treatment led to a dramatic decrease in responding in the OT-DEV group. Although this situation is often referred to as a ‘reinforced test’ to contrast it with extinction testing, it is perhaps more accurate to describe it as a ‘punished test’ since the net effect of this treatment is a decrement in responding. What is most striking about these results is not the sensitivity to outcome devaluation per se, which can be explained by stimulus-response theory, but the rapidity with which performance was suppressed once the devalued outcome was actually delivered (see [36] for a similar result). The stimulus-response account, of course, predicts a gradual decrease in the strength of habitual responding over trials, as the mediating stimulus-response association is weakened (or inhibited). An alternative to this view, however, predicts a much more sudden shift in the sensitivity of instrumental performance to outcome devaluation.

Although certain types of training encourage a transition from goal-directed to habitual performance, this transition need not be permanent. While this topic has received relatively little attention in the free-operant conditioning literature, it is generally recognized that humans can exert cognitive control over their habitual, or automatic, behavior when conditions arise that warrant deliberation, such as when there is a change in task requirements or when the task must be performed in a dangerous situation [46]. Obviously, being punished for performing a response that had once produced a desirable outcome is a prime example of a situation that requires cognitive control. From this perspective, the outcome devaluation effect that emerges in a punished test does not result from a reduction in habit strength (i.e., through stimulus-response learning), but is instead the result of an acute transition in behavioral control; i.e., a shift from habitual to goal-directed control of performance.

Of course, these two accounts are difficult to distinguish on purely behavioral grounds. However, support for the cognitive control hypothesis can be found in the results of brain lesion studies. As mentioned above, discrete lesions of the pDMS, regardless of whether they are made before or after initial training, have been shown to disrupt the sensitivity of instrumental performance to action-outcome contingency degradation and outcome devaluation when tested in extinction [40]. Such findings indicate that the pDMS is a critical part of the neural system mediating goal-directed instrumental action. Without this structure, one assumes, rats must rely on habit learning, even in situations that should encourage the use of a goal-directed strategy. Thus, their performance should provide us with a means to evaluate the characteristics of purely habitual instrumental responding in the absence of cognitive control. Figure 3(b) shows the results of an experiment in which pDMS-lesioned rats were administered a 'punished' outcome devaluation test. As when tested in extinction, their performance was virtually insensitive to outcome devaluation even though they were given immediate feedback about the current values of the training outcomes. This finding suggests that the learning process mediating punishment in the habit system is exceedingly slow (or at least is not engaged very strongly by the contingent delivery of a devalued outcome). It seems unlikely, therefore, that such a process is responsible for the sudden sensitivity to outcome devaluation seen in the habitual performance of normal (unlesioned) rats when given a punished test, and rather suggests that this effect is more likely the result of a shift in strategy to goal-directed action.

In fact, this finding is not unique. In addition to showing impaired sensitivity to outcome devaluation when tested under extinction, the instrumental performance of BLA-lesioned rats is also surprisingly resistant to the suppressive effect of this treatment when the devalued outcome is contingently delivered at test [42]. Importantly, the pDMS and BLA are both implicated specifically in the expression of goal-directed instrumental performance [40,47]. For instance, it has been shown that post-training lesions of these structures are as effective in abolishing the effect of outcome devaluation on extinction performance as pre-training lesions. In contrast, lesions of the PL and MD are effective in disrupting outcome devaluation performance only if made before instrumental training, suggesting that these area play a more restricted role in goal-directed action limited to initial acquisition processes [40,47,48]. In light of these different patterns of involvement, it is interesting that, unlike rats with BLA or pDMS lesions, rats with lesions of PL or MD display normal sensitivity to outcome devaluation when given response-contingent feedback about the current value of the training outcomes [38,41], suggesting that these lesions preserved some capacity for goal-directed instrumental action selection.

Modeling compulsive drug seeking in animals

It is instructive at this point to reconsider the habit learning account of drug addiction in the light of the behavioral findings described above. Indeed, based on these findings, it seems unlikely that drugs produce compulsive behavior merely by facilitating the rate of habit

formation. Although this might address some of the inflexibility of drug-seeking behavior, given what we know about the habitual control of action selection in both rodents and humans this position should anticipate that the addict will suddenly revert to a goal-directed strategy once they begin to experience the adverse consequences of that behavior, which constitute a form of punishment. Therefore, based on the data describe above, a more accurate account would assume that truly compulsive drug seeking is *dominated* by the habit system; i.e. that addicts – as with BLA and pDMS lesioned rats – have difficulty re-exerting goal-directed control over their behavior even in the face of significant negative feedback. This account would propose, therefore, that, not only do drugs of abuse lead to a more powerful engagement of the habit-learning process, they also result in down-regulation of the circuitry mediating goal-directed action through which animals rapidly exert control over habits to suppress their influence on performance when they become maladaptive.

The implications of this position for models of drug seeking are reasonably clear. Perhaps the most straightforward way to model compulsive drug seeking in the laboratory is to train subjects to self-administer a particular drug by performing a response, like lever pressing. Since it involves instrumental conditioning, the drug self-administration paradigm would also seem to lend itself well to attempts to evaluate the role of habit learning in addiction. However, there are a number of methodological problems inherent in this approach. First, unlike natural rewards, it is not clear how one would experimentally devalue a drug reward, particularly when that drug in delivered intravenously. The concept of outcome devaluation, as a method for studying the content of associative learning, assumes that the subject has acquired some cognitive representation of the outcome event that can be distinguished from – and therefore processed independently of – other outcomes. A drug infusion, although producing an internal hedonic effect, tends to lack sufficiently salient local sensory features that would normally compose the representation of a food outcome (i.e., its color, shape, smell, taste [49]).

One way to circumvent this problem is to have the rat voluntarily ingest the drug outcome, although this approach introduces its own problems, as animals often resist consuming these substances which are typically bitter. However, rodents will come to consume certain drugs, like alcohol and cocaine, if they are presented in sweet solution [50–53]. Using this approach, Miles et al. [52] trained rats to perform one action for a lemon-sucrose solution and another for a cocaine-sucrose solution. Between training and testing, each rat had one of these two outcomes devalued through conditioned taste aversion training. Not surprisingly, the group for which the lemon-sucrose outcome was devalued showed a selective suppression of this response, demonstrating that they were able to exert goal-directed control over their behavior. More importantly, however, devaluing the cocaine-sucrose outcome failed to have any effect on performance of the response that had earned this outcome, suggesting that cocaine reinforcement had resulted in accelerated habit formation.

This approach has also been applied to alcohol-reinforced responding with similar results; i.e., rats failed to suppress their performance of an action trained with an alcohol-sucrose solution after that outcome had been devalued [51]. In these studies, the sensitivity of instrumental performance was assessed in extinction. Thus, while they are consist with the habit learning account of addiction, these tests tell us little about whether or not these drug delivery protocols generate truly compulsive behavior, which we argue involves persisting in a habitual response under conditions that should encourage a transition to goal-directed performance. However, in both of these studies, extinction testing was followed by another test in which rats were given the opportunity to earn the valued and non-devalued outcomes by performing the appropriate responses; i.e., they were now ‘punished’ for performing the action that was trained with the devalued outcome [51,52]. Interestingly, although devaluing the drug outcomes (cocaine-sucrose or alcohol-sucrose) did not affect instrumental performance in extinction, in

these studies they were effective in quickly suppressing performance of their respective actions when they were delivered contingently in the punished test.

The fact that, in both cases, rats were able to rapidly regain control over what appeared to be habitual responding indicates that this drug-administration protocol did not produce truly compulsive drug seeking behavior. However, there have been several recent reports that cocaine-reinforcement can produce behavior that is insensitive to response-contingent punishment, which in these studies involved the delivery of either a mild footshock stimulus [54] or a CS that signaled foot shock [55]. In these studies, however, cocaine was delivered intraveneously at doses likely to have a significantly greater pharmacological effect than those used in the oral administration studies. Furthermore, insensitivity to punishment was found to occur only after extensive cocaine self-administration training, consistent with the habit learning account of drug addiction.

Of course, there are alternative explanations for such findings. For example, it is possible that, rather than responding habitually, the rats were in fact responding in a goal-directed fashion, but had come to overvalue the cocaine outcome. From this perspective, one should expect rats to tolerate occasional punishment if it does not outweigh the high incentive value attributed to cocaine. However, this account also predicts that their performance should be sensitive to changes in outcome value and in the action-outcome contingency. Unfortunately, the use of intravenous drug delivery precludes the direct application of the standard tests of these factors in these experiments. Although we are not aware of any studies assessing the effect of cocaine self-administration on instrumental contingency degradation, it has been shown that rats resistant to punishment are also willing to work harder (i.e., perform more responses) for cocaine reinforcement than their punishment insensitive counterparts when given a progressive ratio test. While this finding is hardly conclusive, it could be taken as evidence that these rats were capable of exerting goal-directed control over their actions.

Although using the drug self-administration paradigm to model compulsive drug seeking has some appeal, it also has a number of limitations. We have already noted that this approach does not lend itself to tests of post-training outcome devaluation, which severely constrains attempts to investigate the content of learning in these experiments. In addition, while having a rat self-administer a drug would seem to have considerable face validity as a model of drug seeking, it could be argued that such tasks are better suited for modeling drug *taking* since the behavior being targeted in such studies is proximal to drug delivery. Some researchers have addressed this problem by training rats to perform a heterogeneous chain of two responses, ensuring that the initial link in the chain, the drug *seeking* response, is temporally distal to the drug reward, relative to the terminal, or taking, response [56]. Indeed, there is earlier evidence that, when natural rewards are used to reinforce performance, the distal and proximal actions fall under the influence of distinct motivational processes [28,57]. However, it remains unknown whether this dissociation also applies to drug reinforced response chains. It is possible, for example, that initial and terminal links in the chain need to be isolated from each other even further in order successfully to model the characteristics of both drug seeking and drug taking.

Another problem with the drug self-administration paradigm is that it conflates several learning processes, each of which may contribute to the generation of compulsive behavior. So far we have focused on the two learning processes thought to mediate instrumental action selection. Thus, the question has remained open whether subjects respond because they have learned that an action results in drug delivery (action-outcome learning) or because that action has become associated with contextual cues (stimulus-response learning). However, subjects also have the opportunity to learn about the Pavlovian relationship that exists between these contextual cues and the drug delivery. Although it is often acknowledged that Pavlovian learning may contribute to the compulsive nature of drug seeking [2], using the drug self-administration

approach makes it difficult (and perhaps impossible) to evaluate its role independently of these other learning processes. For example, it is possible that context-drug learning is solely responsible for generating compulsive drug seeking because the mere presence of such cues biases the control of behavior towards the habit system (or away from the goal-directed system). But it is hard to imagine how one could test this account using the self-administration paradigm since manipulations of context should affect the expression of both context-drug and context-response associations.

Finally, the recent growth in research in the neuroscience of decision-making [58,59] has brought with it expansion in related areas most notably into the effect of drug taking on decision making more generally, as opposed to drug taking and drug-related hedonic processes in particular. Establishing the neural bases of drug seeking is, however, made more difficult because of the need to understand two complex interacting factors: (i) the neural processes mediating reward seeking and decision-making on the one hand and (ii) the effect of drug ingestion on those neural processes on the other. This is a particular issue for drug seeking because, in the ordinary course of events, these two processes are hopelessly confounded; humans and other animals self-administering drugs are both engaged in reward seeking – and so, presumably, have activated the neural circuits and systems involved in this activity – and are also ingesting drug, producing any associated changes in the neural processes mediating reward seeking and resulting in compulsive drug seeking. The self-administration approach to studying drug addiction, whilst maintaining a degree of face validity, makes it significantly more difficult to establish whether neural processes engaged during drug pursuit are specific to drug pursuit or are a product of changes in neural processing induced by drug ingestion.

In response to these issues we propose that a more appropriate means to study the influence of drugs on decision-making generally and reward-seeking in particular is to focus on the influence of drug exposure on normal decision processes. To this end, the effects of drug exposure, whether induced by peripheral and intra-cerebral drug administration, when accompanied by behavioral tasks like those described above that allow the assessment of changes in the control of actions by goal-directed and habitual processes and attendant motivational mechanisms would appear to offer a suitable initial approach.

Conclusion

During the development of addiction the pursuit of drugs of abuse becomes progressively less goal-directed and progressively more habitual coming under the control of internal and external states and stimuli. Understandably, therefore, recent theory and research on addiction has begun to focus on the habit learning process and its behavioral and neural bases. It is important, however, to distinguish habitual drug-seeking from other forms of habitual behavior. Under normal conditions, habit learning can be highly adaptive; habits allow us and other animals to relegate the control of routine behavioral responses to a system that uses few cognitive resources freeing up this limited capacity for tasks that need greater monitoring. Unlike goal-directed actions, that are quickly acquired and flexibly deployed, habits are usually slowly acquired, stimulus-bound and inflexible. Nevertheless, their deployment can be rapidly suppressed when conditions change. Driving, cycling, even walking would be very dangerous activities if we couldn't quickly and reliably suppress these habits when circumstances change.

In contrast, habitual drug seeking is pathological; drug exposure appears both to increase the rate of habit acquisition as well as the influence of drug-associated contexts and cues on performance. Furthermore, despite the heavy emphasis on habit processes in current research, a distinguishing feature of habitual drug seeking is the addicts' loss of executive or behavioral control over the habit; drug seeking persists even in the face of severe negative consequences. The compulsive pursuit of drugs can be viewed, therefore, as the product of two interacting

processes: (i) a drug-induced increment in the acquisition of habitual drug seeking and (ii) a drug-induced decrement in the addict's ability to exert control over the habit in the face of persistent, sometimes extreme negative feedback. It is important to recognize, however, that these effects of drug exposure extend beyond drug seeking and effect decision-making and adaptive behavioral control more generally. It is likely, therefore, that the pathological pursuit of drugs reflects changes in the larger neural systems involved in the acquisition and performance of goal-directed and habitual actions.

This account overcomes some of the general problems identified with the purely habit based account of drug addiction. For example, one argument against the claim that drug addiction reflects an abnormal increment in habit learning has been based on, albeit largely anecdotal, evidence of the highly devious and nefarious strategies that addicts devise in procuring drugs. The perspective proposed here sidesteps this kind of issue by emphasizing the pathological nature of the habitual control induced, not simply by an increment in habit learning but by drug-induced abnormalities in the goal-directed system with the consequent changes in goal-directed decision-making processes and in behavioral control.

Acknowledgments

The preparation of this manuscript was supported by a grant from the NIAAA to BWB and SBO: # AA018014

References

1. Everitt BJ, et al. Associative processes in addiction and reward. The role of amygdala-ventral striatal subsystems. Ann N Y Acad Sci 1999;877:412–438. [PubMed: 10415662]
2. Everitt BJ, Robbins TW. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nat Neurosci 2005;8 (11):1481–1489. [PubMed: 16251991]
3. Redish A, et al. A unified framework for addiction: vulnerabilities in the decision process. Behav Brain Sci 2008;31:415–437. [PubMed: 18662461]
4. Tiffany S, et al. What can dependence theories tell us about assessing the emergence of tobacco dependence? Addiction Suppl 2004;1:78–86.
5. Dickinson A. Actions and habits: the development of behavioural autonomy. Philosophical Transactions of the Royal Society of London 1985;B308:67–78.
6. Dickinson, A. Instrumental conditioning. In: Mackintosh, NJ., editor. Animal cognition and learning. Academic Press; 1994. p. 4-79.
7. Adams CD, Dickinson A. Instrumental responding following reinforcer devaluation. Quarterly Journal of Experimental Psychology 1981;33B:109–121.
8. Frese, M.; Sabini, J. Goal directed behavior: The concept of action in psychology. Lawrence Erlbaum Associates; 1985.
9. Dickinson, A.; Balleine, BW. Actions and responses: The dual psychology of behaviour. In: Eilan, N., et al., editors. Spatial Representation. Basil Blackwell Ltd; 1993. p. 277-293.
10. Dickinson A, Mulatero CW. Reinforcer specificity of the suppression of instrumental performance on a non-contingent schedule. Behavioural Processes 1989;19:167–180.
11. Hammond LJ. The effect of contingency upon appetitive conditioning of free operant behavior. Journal of Experimental Analysis of Behavior 1980;34:297–304.
12. Colwill, RM.; Rescorla, RA. Associative structures in instrumental learning. In: Bower, GH., editor. The psychology of learning and motivation. Vol. 20. Academic Press; 1986. p. 55-104.
13. Balleine BW, Dickinson A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 1998;37 (4–5):407–419. [PubMed: 9704982]
14. Davis J, Bitterman ME. Differential reinforcement of other behavior (DRO): A yoked-control comparision. Journal of the Experimental Analysis of Behavior 1971;15:237–241. [PubMed: 16811508]

15. Balleine, BW. Incentive processes in instrumental conditioning. In: Klein, RMS., editor. Handbook of contemporary learning theories. LEA; 2001. p. 307-366.
16. Dickinson A, Balleine BW. Motivational control of goal-directed action. Animal Learning & Behavior 1994;22:1–18.
17. Balleine BW, Dickinson A. The role of incentive learning in instrumental outcome revaluation by specific satiety. Animal Learning & Behavior 1998;26:46–59.
18. Balleine, BW. Taste, disgust and value: Taste aversion learning and outcome encoding in instrumental conditioning. In: Reilly, S.; TRS, editors. Conditioned Taste Aversion: Behavioral and Neural Processes. Oxford University Press; 2009. p. 262-280.
19. Balleine BW, Ostlund SB. Still at the choice point: Action selection and initiation in instrumental conditioning. Annals of the New York Academy of Sciences 2007;1104:147–171. [PubMed: 17360797]
20. Rescorla RA. Response-outcome versus outcome-response associations in instrumental learning. Animal Learning & Behavior 1992;20:223–232.
21. Rescorla RA, Solomon RL. Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. Psychol Rev 1967;74 (3):151–182. [PubMed: 5342881]
22. Trapold, MA.; Overmier, JB. The second learning process in instrumental conditioning. In: Black, AA.; Prokasy, WF., editors. Classical Conditioning: II. Current research and theory. Appleton-Century-Crofts; 1972. p. 427-452.
23. Corbit LH, et al. The role of the nucleus accumbens in instrumental conditioning: Evidence of a functional dissociation between accumbens core and shell. J Neurosci 2001;21 (9):3251–3260. [PubMed: 11312310]
24. Dickinson A, et al. Bidirectional instrumental conditioning. Q J Exp Psychol B 1996;49 (4):289–306. [PubMed: 8962537]
25. Robbins TW, et al. Limbic-striatal interactions in reward-related processes. Neurosci Biobehav Rev 1989;13 (2–3):155–162. [PubMed: 2682402]
26. Winterbauer, NE. Unpublished PHD Dissertation. 2006. Conditioned reinforcment.
27. Dickinson, A.; Balleine, BW. The role of learning in the operation of motivational systems. In: Gallistel, CR., editor. Learning, Motivation & Emotion, Volume 3 of Steven's Handbook of Experimental Psychology. 3. John Wiley & Sons; 2002. p. 497-533.
28. Corbit LH, Balleine BW. Instrumental and Pavlovian incentive processes have dissociable effects on components of a heterogeneous instrumental chain. J Exp Psychol Anim Behav Process 2003;29 (2): 99–106. [PubMed: 12735274]
29. Corbit LH, Balleine BW. Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. J Neurosci 2005;25 (4):962–970. [PubMed: 15673677]
30. Kruse JM, et al. Pavlovian conditioned stimulus effects upon instrumental choice behavior are reinforcer specific. Learning and Motivation 1983;14:165–181.
31. Corbit LH, et al. General and outcome-specific forms of Pavlovian-instrumental transfer: The effect of shifts in motivational state and inactivation of the ventral tegmental area. European Journal of Neuroscience 2007;26:3141–3149. [PubMed: 18005062]
32. Holland PC. Relations between Pavlovian-instrumental transfer and reinforcer devaluation. J Exp Psychol Anim Behav Process 2004;30 (2):104–117. [PubMed: 15078120]
33. Rescorla RA. Transfer of instrumental control mediated by a devalued outcome. Animal Learning & Behavior 1994;22:27–33.
34. Balleine BW. Instrumental performance following a shift in primary motivation depends on incentive learning. J Exp Psychol Anim Behav Process 1992;18 (3):236–250. [PubMed: 1619392]
35. Dickinson A, et al. The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. Quarterly Journal of Experimental Psychology 1983;35B:35–51.
36. Adams CD. Variations in the sensitivity of instrumental responding to reinforcer devalaution. Quarterly Journal of Experimental Psychology 1981;34B:77–98.
37. Dickinson A, et al. Omission learning after instrumental pretraining. Quarterly Journal of Experimental Psychology 1998;51B:271–286.

38. Corbit LH, Balleine BW. The role of prelimbic cortex in instrumental conditioning. Behav Brain Res 2003;146 (1–2):145–157. [PubMed: 14643467]
39. Yin HH, et al. Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. Eur J Neurosci 2005;22 (2):505–512. [PubMed: 16045503]
40. Yin HH, et al. The role of the dorsomedial striatum in instrumental conditioning. Eur J Neurosci 2005;22 (2):513–523. [PubMed: 16045504]
41. Corbit LH, et al. Lesions of mediodorsal thalamus and anterior thalamic nuclei produce dissociable effects on instrumental conditioning in rats. Eur J Neurosci 2003;18 (5):1286–1294. [PubMed: 12956727]
42. Balleine BW, et al. The effect of lesions of the basolateral amygdala on instrumental conditioning. J Neurosci 2003;23 (2):666–675. [PubMed: 12533626]
43. Yin HH, et al. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur J Neurosci 2004;19 (1):181–189. [PubMed: 14750976]
44. Yin HH, et al. Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. Behav Brain Res. 2005
45. Killcross S, Coutureau E. Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb Cortex 2003;13 (4):400–408. [PubMed: 12631569]
46. Norman, D.; Shallice, T. Vol. Technical Report #99. Centre for Human Information Processing, University of California; San Diego: 1980. Attention to action: Willed and automatic control of behaviour.
47. Ostlund SB, Balleine BW. Differential involvement of the basolateral amygdala and mediodorsal thalamus in instrumental action selection. Journal of Neuroscience 2008;28:4398–4405. [PubMed: 18434518]
48. Ostlund SB, Balleine BW. Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. J Neurosci 2005;25 (34):7763–7770. [PubMed: 16120777]
49. Pelloux Y, et al. Compulsive drug seeking by rats under punishment: effects of drug taking history. Psychopharmacology (Berl) 2007;194:127–137. [PubMed: 17514480]
50. Corbit L, Janak P. Ethanol-associated cues produce general pavlovian-instrumental transfer. Alcohol Clin Exp Res 2007;31:766–774. [PubMed: 17378919]
51. Dickinson A, et al. Alcohol seeking by rats: action or habit? Q J Exp Psychol B 2002;55:331–348. [PubMed: 12350285]
52. Miles F, et al. Oral cocaine seeking by rats: action or habit? Behav Neurosci 2003;117:927–938. [PubMed: 14570543]
53. Samson H, Doyle T. Oral ethanol self-administration in the rat: effect of naloxone. Pharmacol Biochem Behav 1985;22:91–99. [PubMed: 3975250]
54. Deroche-Gamonet V, et al. Evidence for addiction-like behavior in the rat. Science 2004;305:1014–1017. [PubMed: 15310906]
55. Vanderschuren L, Everitt B. Drug seeking becomes compulsive after prolonged cocaine self-administration. Science 2004;305:1017–1019. [PubMed: 15310907]
56. Olmstead M, et al. Cocaine seeking by rats is a goal-directed action. Behav Neurosci 2001;115:394–402. [PubMed: 11345964]
57. Balleine BW, et al. Motivational control of heterogeneous instrumental chains. Journal of Experimental Psychology: Animal Behavior Processes 1995;21:203–217.
58. Balleine BW, et al. Current trends in decision making. Ann N Y Acad Sci 2007;1104:xi–xv. [PubMed: 17595291]
59. Balleine, BW., et al., editors. Reward and Decision Making in Corticobasal Ganglia Networks. New York Academy of Sciences; 2007.
60. Bolles RC, Holtz R, Dunn T, Hill W. Comparisons of stimulus learning and response learning in a punishment situation. Learning and Motivation 1980;11:78–96.

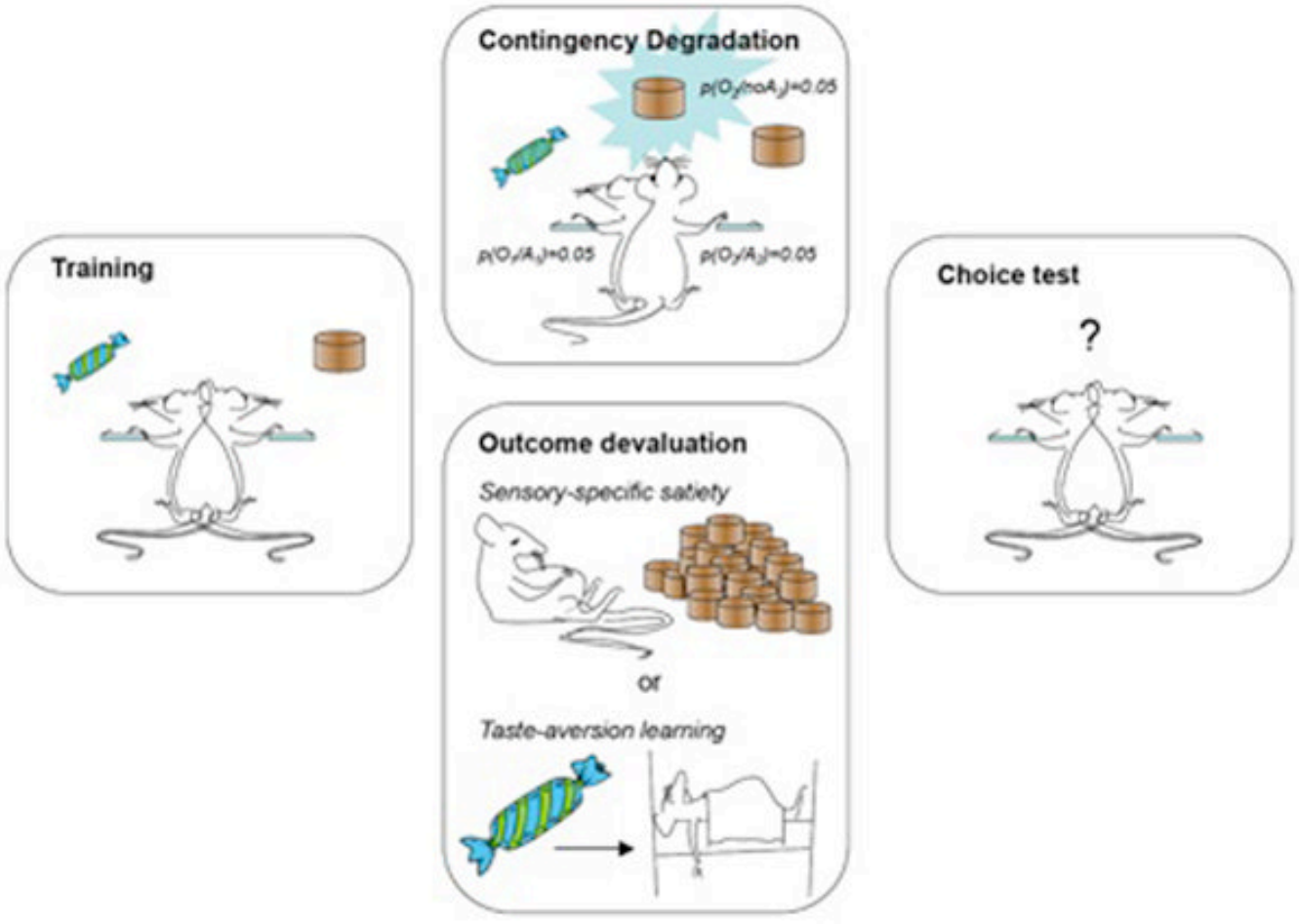

Figure 1.

Illustration of contingency degradation and outcome devaluation procedures used to test whether actions are goal-directed in instrumental conditioning. Rats are trained on two actions, e.g. two levers, and rewarded with different outcomes for each action, here food pellets and sugar (left panel). After training two kinds of test are conducted: (i) The contingency degradation test (depicted in the top-center panel) in which one or other outcome is delivered non-contingently at the same probability as it is earned contingent on lever pressing. The other outcome continued to be earned only by lever pressing. (ii) An outcome devaluation test (depicted in the lower-center panel) prior to which one or other outcome is devalued either by sensory-specific satiety (top) or taste aversion learning (bottom). After each of these treatments rats are given a choice test in extinction (right panel) to assess the effects of the degradation and devaluation manipulations on choice.

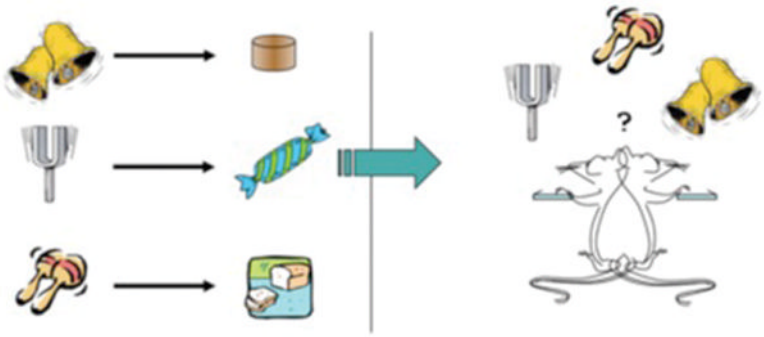

Figure 2.

Pavlovian-instrumental transfer. In this depiction, rats are first given pairings between three different auditory stimuli and three different food outcomes. Later they are trained on two actions, say two levers, delivering two of the three outcomes used in the first phase as depicted in the left panel of Figure 1. They are subsequently given a choice test on the two levers in which the three stimuli are presented in extinction. As previously reported (e.g. [29], [31]), this treatment generates evidence for two forms of transfer effect: (i) a general form, which is here depicted by the ‘maracas stimulus paired with a food outcome that is not then earned by lever pressing the presentation of which results in a general increase in the performance of both actions and (ii) a specific form, here depicted by the ‘bell’ and ‘tuning fork’ stimuli associated with the food outcomes that were also earned by pressing the levers. The effect of the stimulus presentation in this situation is to bias choice towards the action that, in training, earned the outcome predicted by the stimulus; e.g. if the left lever earned sugar then the tuning fork would enhance responding on that lever and not the bell, which would bias responding towards the right lever.

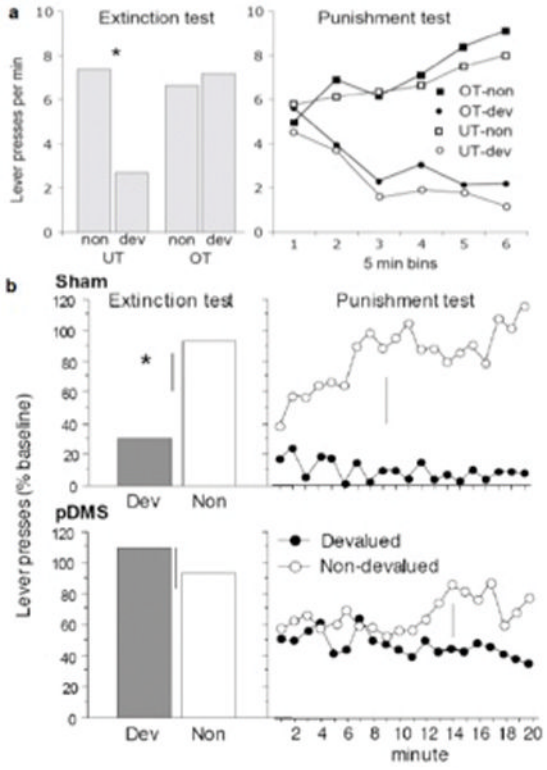

Figure 3.

(a) Results of a devaluation extinction test (left panel) and punishment test (right panel) conducted after rats have been given undertrained (UT) or overtrained (OT) to lever press for sugar and then the sugar devalued by taste aversion learning. (b) Results from a choice outcome devaluation extinction test (left panels) and punishment test (right panels) in which rats have been trained to press two levers for distinct outcomes after being given either sham surgery or bilateral lesions of posterior dorsomedial striatum (pDMS). See text for details.