Published in final edited form as:
Behav Processes. 2012 May ; 90(1): 9–19. doi:10.1016/j.beproc.2012.03.006.

Issues in the Extinction of Specific Stimulus-Outcome Associations in Pavlovian Conditioning

Andrew R. Delamater
Brooklyn College of the City University of New York

Abstract

This paper reviews a variety of studies designed to examine the effects of extinction upon control by specific stimulus-outcome (S-O) associations in Pavlovian conditioning. Studies conducted with rats in a magazine approach conditioning paradigm have shown that control by specific S-O associations is normally unaffected by extinction treatments, although other aspects of conditioned responding seem affected in a more enduring way. However, recent work suggests that extinction can undermine control by such associations if it is administered after the conditioned stimulus is weakly encoded. The results from these studies suggest that it may be important to consider multiple response systems in assessing the impact of extinction. Studies conducted with the flavor preference learning paradigm in rats also show that specific S-O associations can be undermined by procedures that involve presenting a flavor cue in the absence of its associated nutrient. These findings provide no support for the view that flavor preference learning necessarily entails some unique learning process that differs from more conventional processes. As in other situations, some of these effects likely involve a masking process, but the extent to which masking or true associative weakening occurs in extinction more generally is a topic that is not well understood. Finally, we present some data to suggest that extinction also involves conditional "occasion-setting" control by contextual cues. Special procedures are recommended in assessing such learning when the goal is to distinguish this form of learning from other more conventional mechanisms of extinction.

Keywords

Magazine approach; Flavor Preference Learning; Contextual Occasion Setting

The study of extinction has experienced a renaissance in recent years. Much of this has to do with the rapid progress being made in the identification of the neural substrates of basic acquisition and extinction processes (e.g., Myers & Davis, 2007; Maren & Quirk, 2004; Quirk & Mueller, 2008; Sotres-Bayon, Bush, & LeDoux, 2004). Particularly exciting are those developments in which basic neural and psychological mechanisms are intersecting in the analysis of extinction (e.g., see Delamater, 2004). One clear example of this is the study of context-dependent extinction processes and the modulatory role of the hippocampus (e.g., see Bouton, Westbrook, Corcoran, & Maren, 2006; Maren & Quirk, 2004). In spite of these success stories, though, there are a large number of basic issues (both of a psychological and

Address Correspondence to: Andrew R. Delamater Psychology Department Brooklyn College – CUNY 2900 Bedford Ave Brooklyn, NY 11210 USA, andrewd@brooklyn.cuny.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors maybe discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

neural systems nature) that require further investigation. For example, my own research on extinction over the years has focused on an aspect of the phenomenon that has received very little attention in the literature. In particular, I have chosen tasks that permit for an identification of some very specific aspect of learning – sensory-specific stimulus-outcome associations – and asked whether extinction treatments might have some impact on control by that particular aspect of learning. Considering extinction from this perspective opens the field to additional questions concerning the nature of extinction effects on other aspects of learning as well, whereby different processes may be differentially affected. What follows is a review of the key findings, mostly from my lab, concerning the nature of extinction effects mostly on this aspect of learning.

My starting point is with the now classic Rescorla-Wagner model of Pavlovian conditioning (Rescorla & Wagner, 1972). The reason for starting here is because this model also, in some ways, has experienced a renaissance of a sort. As the neural mechanisms of basic learning processes are being uncovered it is becoming clearer that the precise mechanisms identified by Rescorla and Wagner are providing a reasonably accurate depiction of the basic mechanisms driving neural plasticity. The time is right to consider, once again, how this model handles basic phenomena of extinction. Another reason why it is worthwhile to consider these ideas again in the study of extinction is that it is frequently the case that investigators fail to consider the involvement of basic processes in lieu of other more complex mechanisms. It could very well turn out that both types of mechanisms (to be identified shortly) play a role in extinction, and, if so, it would be well advised for us to consider this in greater detail.

Revisiting the Rescorla-Wagner Model

The Rescorla-Wagner model (Rescorla & Wagner, 1972) continues to be one of the dominant theoretical approaches to the study of simple associative learning processes, particularly of Pavlovian learning. According to this model, learning is construed in terms of changes in the associative (or “connection”) strength between representations of conditioned and unconditioned stimuli, CS and US, respectively, that take place when these two events are optimally paired. To be sure, there remain problems with some of the model’s specific assumptions (e.g., Gallistel & Gibbon, 2000; Miller, Barnet, & Grahame, 1995; Pearce, 2002), but, at the same time, some of the key elements of the model appear to accurately capture the way in which real nervous systems learn.

One of these ideas is that the basic driving force behind learning is what has commonly come to be referred to as a “prediction error.” Such errors occur when the organism generates an expectancy of the US, on the basis of a particular CS in the environment, that does not accurately reflect the current state of affairs. For instance, early in training before a CS has had much opportunity to associate with a US, the CS will generate only a very weak expectation of the US (or none at all). When the US actually occurs on the conditioning trial, then its occurrence is said to be surprising because it was not adequately anticipated. That is, a positive prediction error occurs. As conditioning proceeds, however, the associative strength between the CS and US increases, ultimately, to some asymptotic level, at which point the CS generates a strong expectancy of the US’s occurrence. Under these circumstances, when the US does occur, it is, in a sense, fully anticipated by the CS and no surprise takes place. In other words, no prediction error is generated. No further learning will take place on these conditioning trials.

In contrast, after learning has occurred a different kind of prediction error is generated on extinction trials. When a fully trained CS is presented without the US during extinction, for instance, then a fully anticipated US fails to occur. This generates a negative prediction error and according to the model will cause a negative change in associative strength. In other words, the excitatory association that had been established over the course of the conditioning phase will now become weakened, until, ultimately, extinction might come to fully abolish this excitatory association.

Current investigations of the neural mechanisms of simple Pavlovian conditioning have revealed impressive support for the basic idea that positive and negative prediction errors are coded by either individual neurons or neural circuits and are responsible for neural plasticity within the system. For instance, Schultz and his colleagues (e.g., Tobler, Dickinson, & Schultz, 2003; Waelti, Dickinson, & Schultz, 2001) have found evidence to suggest that individual mid-brain dopamine cells respond by increasing their firing rate when unanticipated, but not anticipated, juice USs are presented. The same stimuli respond by decreasing their firing rate when the anticipated juice US is omitted on an extinction trial. In addition, authors have also provided more direct evidence for the existence of conditioning-dependent negative feedback circuits that limit the effective processing of USs when they are anticipated by the CS (e.g., see Kim, Krupa, & Thompson, 1998; McNally, Johansen, & Blair, 2011; Schoenbaum, Roesch, Stalnaker, & Takahashi, 2009). While the evidence is increasing to suggest that prediction errors drive conditioning, there are surely elements of the Rescorla-Wagner model that have been less well supported.

The Challenge of Extinction

One of the most basic assumptions of the Rescorla-Wagner model that has received perhaps the most challenge is its so-called “independence of path” assumption (see Bouton, 1991; Pearce & Hall, 1980). According to this assumption the status of a stimulus is given by its current associative value, and it does not matter how such a stimulus may have achieved that value. Consider, on the one hand, a CS that has been conditioned and then extinguished to a moderate level, say, to half of its maximal value. Compare this cue with another CS that has been only incompletely conditioned to the same level (half of its maximal value). These two stimuli should be functionally equivalent because their own associative histories should not matter; their own “paths” to this comparable associative level should be immaterial. So, for example, when authors have found that extinguished CSs behave differently from partially reinforced (Bouton & King, 1986) or latently inhibited CSs (Bouton & King, 1986; Swartzentruber & Rescorla, 1994) trained to the same associative level, such results present strict challenges to this basic assumption of the model.

Another basic challenge, at least at first glance, comes from studies demonstrating that extinction fails to completely erase learning. Phenomena such as spontaneous recovery, reinstatement, renewal, and rapid reacquisition all speak to the fact that at least some of the original learning is preserved following an extinction treatment. However, it is important to note that while each of these phenomena establish that extinction fails to completely abolish previous learning, it is more difficult to use these results to imply that extinction fails to weaken previously learned associations as is anticipated by the Rescorla-Wagner model. The basic problem can be appreciated by considering how the Rescorla-Wagner model explains these phenomena, and I will illustrate by considering its application to the renewal phenomenon.

In the most robust form of renewal a CS is first trained in one context (A), extinguished in a second (B), and then tested either in the acquisition or extinction context. Greater responding is seen to the CS when it is tested in its acquisition compared to its extinction context. This is the so-called ABA renewal effect. According to the Rescorla-Wagner model, this result can occur for two reasons. First, when the CS is extinguished in the presence of context B, then this context acquires inhibitory associative strength because it occurs at a time when nonreinforcement of the CS results in a negative prediction error and because its associative value starts at zero. This inhibitory learning, in turn, will lower the degree to which the US is expected when the CS subsequently occurs because the inhibitory strength of the context will sum with the somewhat reduced excitatory strength possessed by the CS. At the end of extinction, the inhibitory context-US association will, therefore, protect the CS from undergoing complete associative loss – a phenomenon known as “protection from extinction” (e.g., Rescorla, 2003). When the CS is tested outside of the extinction context, then increased levels of responding will be seen to the CS because it will now be tested in a context that is not inhibitory. The remaining residual levels of excitatory strength will become manifest.

The second possible source of renewal arises from the possibility that the original training context has acquired some excitatory associative strength of its own. When the CS is tested back in this context, then the residual excitatory strength of the CS can sum its excitatory effects with the excitatory strength conditioned to the context to result in a stronger CR than would occur when the CS is tested in a more neutral context. Investigators have, indeed, found greater levels of responding to a CS when this is tested in a more excitatory context compared to a less excitatory context (e.g., Brandon & Wagner, 1991; Grau & Rescorla, 1984).

The upshot of this analysis is that according to the Rescorla-Wagner model extinction should rarely, if ever, produce a complete loss in associative strength to the CS because of this protection from extinction mechanism. Further, summation effects (with inhibitory or excitatory contexts) could partly explain some of the findings. While early attempts failed to provide direct evidence for the extinction context acquiring inhibitory strength (Bouton & King, 1983; Bouton & Swartzentruber, 1986), more recent evidence suggests that the extinction context can in fact become inhibitory (Polack, Laborda, & Miller, 2011) and more recent theorizing offers reasons for why such effects may become difficult to observe (Laurrari & Schmajuk, 2008). This analysis will not explain all instances of renewal (see Bouton & King, 1986), but it will be important to keep these mechanisms in mind when attempting to interpret renewal phenomena because they may contribute to renewal in many situations.

That the model anticipates extinction to only partially weaken the associative strength to a CS is often overlooked. Furthermore, the mechanisms noted above may also be used to explain the other key phenomena that are often used to demonstrate that extinction fails to fully undermine learning – namely, spontaneous recovery and US reinstatement. Each of these other phenomena can either be reduced to special cases of renewal (by temporal contexts in spontaneous recovery (see Bouton, 2004; Bouton, et al., 2006)) or context-CS summation (in the case of reinstatement). Thus, demonstrations of spontaneous recovery, renewal, reinstatement, or even fast reacquisition are all entirely consistent with predictions derived from the Rescorla Wagner model, and should not, in and of themselves, be seen as inconsistent with this model. One qualification here, however, is that context-CS summation has sometimes been shown to work more effectively with trained and then extinguished CSs than with nonextinguished CSs. This would imply a more complex interaction than would be anticipated by the Rescorla-Wagner model (e.g., see Bouton, 1991). Such observations, however, should not lead us to conclude that summation processes never play a role in renewal-related phenomena.

Does Extinction Weaken the Underlying Association?

However, there are additional data that question the basic notion that extinction should result in any weakening at all of the associative strength possessed by a CS. Both Delamater

(1996) and Rescorla (1996) compared the status of the associative strengths to a CS that had undergone extinction to one that had not. In both cases, the levels of associative strength to these two stimuli were equivalent. The Delamater (1996) study is particularly instructive because in that study there were two measures of responding, one of which was affected by extinction and one of which was not. This study used a Pavlovian to instrumental transfer test (PIT) to assess the status of the association formed between the CS and the specific sensory properties of the US. Initially, two different instrumental responses were trained in separate sessions with variable interval schedules of reinforcement (e.g., Lever press – Pellets, Chain pull – Sucrose). Then, Pavlovian conditioning with two separate CS-US pairs was conducted off-baseline (e.g., Tone – Pellets, Light – Sucrose). Only one of these CSs then underwent extinction over the next 10 sessions (e.g., Tone −). Finally, in a transfer test the rats chose between the two instrumental responses under extinction conditions in the presence of either the Tone or Light CSs, or in the absence of any stimuli. Reduced levels of magazine approach responses recorded during the CSs in this transfer test session were seen to the extinguished CS compared to the non-extinguished CS. However, both of the CSs exerted an outcome-specific increase in the instrumental response with which they shared a reinforcing outcome. For instance, in the presence of the Light CS the rats increased their rate of Chain pull responding over pre CS baseline levels but did not change their rate of Lever press responding. Importantly, the extinguished CS exerted the same degree of selective transfer.

This selective PIT effect has often been interpreted as reflecting the learning of a CS-US association that is quite specific in its sensory content (e.g., Delamater & Holland, 2008; Kruse, Overmier, Konz, & Rokke, 1983). Only if the CS had associated with the specific sensory properties of the US would it be capable of selectively biasing instrumental choice behavior. What this experiment revealed was that extinction did not weaken this specific CS-US association. Extinction was not entirely without effect, however, because in the same animals at the same time the extinguished CS did effectively diminish magazine approach responses. One might argue that the transfer measure was merely less sensitive at detecting changes in associative value that may have occurred with extinction. This possibility is unlikely because (1) we have other data to suggest that learning can be revealed with the transfer measure before being manifested in magazine approach CRs (e.g., Delamater & Oakeshott, 2007), and (2) other research demonstrates that the measure can reveal differences in situations thought to arise when changes in associative value should occur (e.g., Delamater, 1995; Rescorla, 2001). Thus, it appears as though extinction might be having different effects on different aspects of learning.

It is widely accepted now that USs are complex events consisting of multiple components – sensory, emotional, response, temporal, hedonic (e.g., Delamater & Oakeshott, 2007; Konorski, 1967; Wagner & Brandon, 1989). It is possible that when we speak of a CS associating with a US that in fact multiple associative systems are recruited and participate in the learning. Figure 1 illustrates this possibility by showing that a CS might come to form separate associations with each of these different US components. This is not the place to discuss the evidence that bears on this framework (see Delamater, 2012), but rather I would like to suggest that within this framework it seems highly likely that extinction might be expected to have different effects on different components of learning. If each of these learning processes is, to some degree, independent of one another, there would be nothing to prevent an extinction process from hindering one process without influencing another. In this way it seems possible that some associations may be affected (i.e., weakened) by extinction while others are not so affected. When we speak of a CS-US association being established and describe that in terms of some degree of associative strength, it becomes important to realize that there may be many different associative systems that we need to consider because each of these may have their own rules for associative change.

Effect of Memory Strength on Sensitivity to Extinction

Recently, we have been exploring the sensitivity of sensory-specific CS-US associations to extinction when the initial level of conditioning is varied. In a classic human eyeblink conditioning study by Spence, Rutledge, & Talbott (1963), extinction was shown to be more rapid when it occurred following 32 compared to 64 conditioning trials, even though the level of eyeblink responding did not differ at the beginning of extinction testing. More recently, Dudai and colleagues introduced a memory strength hypothesis that could help explain why extinction might be more successful when it is given after relatively few conditioning trials (Eisenberg, Kobilo, Berman, & Dudai, 2003). These authors speculated that the acquisition memory is made stronger and more accessible by more training trials. They further suggested that during an extinction trial, two processes are engaged. First, the CS can retrieve the memory of acquisition and this could induce memory reconsolidation (e.g., Nader, Schafe, & LeDoux, 2000). This process would further strengthen the acquisition memory. Second, because the US does not occur during extinction, this could induce the consolidation of new inhibitory learning (reflecting the fact that the CS occurs without the US). This consolidation of extinction process would antagonize conditioned responding. Finally, they assumed that when extinction occurs after a strong CS-US memory has been established (by many conditioning trials) the CS more actively engages the reconsolidation of acquisition process. Conversely, they assume the CS will more actively engage the consolidation of extinction process when extinction occurs following the establishment of a weak CS-US memory.

There are, of course, other explanations of these sorts of findings. For instance, suppose that following extensive training the associability of the CS has been reduced more than it has following more limited training (Pearce & Hall, 1980). If so, then entering into the extinction phase an extensively trained CS would be more resistant than a CS given more limited training to new inhibitory learning during extinction. In spite of the rather contentious nature of this issue, the fact remains that there is not an extensive amount of research directed to the empirical question of whether the amount of training given to a CS (i.e., memory strength) has any impact on its sensitivity to an extinction treatment.

If extinction were to occur after a more limited number of training trials it might more successfully result in diminished control by or even truly weakened sensory-specific CS-US associations. We have recently conducted experiments in which we assessed the development of sensory-specific CS-US associations as measured by the selective PIT task described above (see Delamater & Oakeshott, 2007). Different groups of rats were given different amounts of Pavlovian training trials before assessing the status of their reinforcer-specific learning in the PIT tests. We observed that after as few as 16 Pavlovian training trials, subjects displayed reward selective PIT reflecting that they had acquired sensory-specific CS-US associations. This number of training trials, however, was considerably below the number of training trials that produce asymptotic levels of conditioned responding using a more conventional conditioned magazine approach response.

In two additional studies we trained rats in this task by administering 16 Pavlovian acquisition trials with each of two CS-US pairs before administering 40 extinction trials spread over 10 sessions. Throughout training and extinction rats were presented with 4 trials of each CS in each session. During training each CS was paired with its respective US, but during extinction sessions no USs were presented. A control group received the same number of training sessions in each study, but during extinction they were placed in the experimental chambers without any CS or US presentations. A PIT test was then administered whereby equal periods in which a CS was presented alternated with periods in which neither CS was presented. Subjects had access to both instrumental responses (lever press and chain pull) but no reinforcers could be earned in this test.

The two studies differed in the manner in which Pavlovian conditioning was conducted. In the first, each CS (noise, flashing light) was presented for 2 min and the USs (pellets, .1 ml of 20% sucrose) were presented 20 s after stimulus onset. The second study was conducted exactly the same way except during Pavlovian training the CS durations were 1 min and the USs occurred at stimulus offset.

In both cases we observed that extinction either totally eliminated or diminished control by sensory-specific associations in the PIT test. Figure 2 displays the results from the PIT test for the extinguished and non-extinguished groups in each study. Displayed is the mean instrumental response rate occurring during the pre-CS baseline periods and also during the CSs. The data (during the CSs) are segregated in terms of the response that was reinforced by the same or different US as that signaled by the CS. More “same” responses than “different” responses indicates that the CS evokes a specific US representation that, in turn, selectively controls the instrumental response that was also reinforced by that same US. In both studies, the non-extinguished control groups displayed selective PIT, but this effect was eliminated in the first study (left) and diminished in the second study (right). The data were analyzed using a Response (pre-CS, same, diff) x Group (extinction, no extinction) ANOVA. A significant Response x Group interaction was found in both Study 1, F(2,58) = 3.72, and Study 2, F(2,120) = 3.65. These interactions were followed by post-hoc tests (Rodger, 1974) showing in the first study that the CSs significantly elevated same and reduced different responses relative to baseline in the non-extinguished control group. The extinguished group showed no effect of the CSs on instrumental responding in this case. In the second study, post-hoc tests revealed that the CSs exerted selective control over responding in both non-extinguished and extinguished groups, but that this control was stronger in the group given no extinction. In both cases, there was more same than different responses made in the presence of the CSs although the effect occurred due to a selective reduction in different responses relative to the pre-CS baseline in the nonextinguished group.

The results of these experiments demonstrate for the first time that extinction impairs selective CS-US associations. Earlier work used either selective PIT (Delamater, 1996) or selective US devaluation tasks (Rescorla, 1996) to study the effects of extinction on the prior learning of specific CS-US associations, and these studies provided no evidence that the associations were weakened in any way. However, in both of those studies considerably more Pavlovian training trials were given prior to extinction than in the studies reported here, and this suggests that the amount of training may be critical in determining the effectiveness of an extinction procedure on control by sensory-specific associations. The results are generally consistent with the hypothesis put forth by Dudai and his colleagues (Eisenberg, et al., 2003), in that weaker memories of acquisition may be especially sensitive to disruption by extinction learning. Additional studies will be required, however, to further support and elaborate these findings (cf, Pearce & Hall, 1980), but, nevertheless, the results demonstrate a clear effect of extinction on sensory-specific CS-US associations. Whether or not this reflects true associative weakening, for example, as presumed by the Rescorla-Wagner model, or masking of such associations will require additional research.

Effects of Non-reinforcement on Sensory-Specific Associations in Flavor Preference Learning

One very popular learning paradigm used to study basic learning processes as well as basic processes of ingestive behaviors is flavor preference conditioning. In one common variation of this paradigm a relatively neutral flavor CS+ (e.g., almond) is paired with (i.e., often mixed in solution with) a nutrient US (e.g., 10% sucrose), while a second flavor (CS−) is presented without the nutrient US. Following such training, rats are given a choice between these two flavors (CS+ vs CS−) in the absence of the nutrient US, and it is often observed that subjects prefer the CS+ flavor to the CS− flavor. This demonstrates that preferences can be conditioned as a result of the flavor-nutrient pairings. The basic effect has been demonstrated in animals trained and tested hungry or thirsty or both. What is less clear from the basic demonstration of a conditioned flavor preference is what particular associations involving the flavor CS and the nutrient US are acquired and contribute to an intake preference.

There are three possible associations noted in the literature. First, the flavor cue might associate with the specific sensory qualities of the nutrient (e.g., the sweet taste of sucrose). If the flavor cue associatively activates a representation of the taste of the nutrient and this taste were palatable, then the animal will consume more of the flavor cue. A second possibility is that the flavor cue might come to associate directly with the positive hedonic response that occurs to the nutrient US. In this case, the flavor CS is preferred not because it activates a representation of the palatable nutrient taste, but, rather, because it becomes capable of evoking a new positive hedonic reaction in and of itself. A third possibility is that the flavor cue becomes preferred because it associates with some post-ingestive reinforcing signal generated by post-oral processing of the nutrient. Some authors have referred to this as a flavor-calorie association.

Of these different types of associations that could mediate preference learning, the third type has been most extensively examined. For example, Sclafani and his colleagues have demonstrated that robust preferences can be established under a variety of conditions when the animal consumes a flavor cue by mouth and receives infusions of a nutrient directly into its stomach (e.g., Elizalde & Sclafani, 1990). In this case, the learned preference can only be of the third type because there is no oral component of the nutrient with which the flavor CS can associate.

While flavor preference learning based upon associations of this last type is well established and much is known about its basic properties, learning of the first two types of associations have been less frequently dissociated. For example, when the flavor CS is paired with a palatable nutrient US by mouth (e.g., sucrose), then how might one distinguish between learning involving the palatable taste of the nutrient or the positive hedonic reaction to that palatable taste? This problem has been addressed by using a US devaluation procedure to pinpoint the presence of specific flavor-taste associations.

For example, suppose that following the conditioning phase the sweet taste of sucrose was rendered unpalatable through its separate pairings with a nausea-inducing agent. Following this aversion conditioning, one can assess the impact this has on the preference for the flavor associate of the now devalued nutrient. If the flavor cue had become preferred because it directly acquired the capacity to evoke its own positive hedonic response, independent of the taste of the nutrient, then devaluing that taste should have no effect on preference for the flavor cue. This follows from demonstrations that hedonic responses are not specifically tied to particular stimuli, but are rather generally activated by a wide range of stimuli within a given hedonic class (e.g., Berridge & Grill, 1984). However, if the flavor cue evokes a representation of the taste of that nutrient and this taste has been devalued prior to the test, then the flavor cue should be avoided as well. A variety of studies have, indeed, shown that acquired flavor preferences are diminished by devaluation of the nutrient US, and this provides the best evidence that specific flavor-taste associations have been formed in this procedure (e.g., Delamater, Campese, LoLordo, & Sclafani, 2006; Dwyer, 2005). What is less clear is how one might best provide evidence of the other type of association indicated above, namely, that between the flavor cue and the hedonic response to the nutrient US. One could infer such learning to the extent that devaluation of the nutrient has no impact on an established preference under certain circumstances (so long as flavor-calorie learning can be ruled out). Another strategy would be to block the formation of specific flavor-taste associations through targeted brain lesion manipulations and then look for preference learning in sham-feeding rats where post-ingestive effects of nutrients are minimized. The problem with this strategy is that it is not yet clear what brain regions are critical in the establishment of specific flavor-taste associations in these paradigms (e.g., see Blundell, Hall, & Killcross, 2003).

We have conducted a number of studies in which we have used the US devaluation task to target learning about the specific flavor-taste association formed during normal flavor preference conditioning procedures. In particular, we have been keenly interested in understanding the effects of non-reinforcement of the flavor cue in this task. One of the commonly-cited features of flavor preference learning is that once established it is difficult to undermine already learned preferences (e.g., Drucker, Ackroff, & Sclafani, 1994; Harris, Shand, Carroll, & Westbrook, 2004). This has prompted some to note that there may be something quite special or unique to learning in this paradigm (e.g., De Houwer, Thomas, & Baeyens, 2001; Pearce, 2002). For instance, De Houwer, et al. (2001) have suggested that flavor preference learning represents a form of “evaluative” conditioning – a type of conditioning process that is qualitatively distinct from more traditional forms of Pavlovian learning. I have noted above how control by sensory-specific CS-US associations in the more traditional magazine approach conditioning paradigm does not appear to be easily undermined (e.g., Delamater, 1996; Rescorla, 1996). If learned preferences for a flavor cue paired with a palatable tasting nutrient were difficult to extinguish because the specific flavor-taste association was relatively insensitive to extinction, then this finding would not be unprecedented, and, indeed, would be entirely consistent with work in the more traditional Pavlovian magazine approach paradigm. Therefore, we set out to explore the effects of extinction upon specific flavor-taste learning in a preference conditioning paradigm.

Delamater (2007) conducted several studies designed to examine this question in rats given flavor preference training and testing while thirsty or while hungry. The experimental design used in two of the studies is illustrated in Figure 3 (top). Initially, two flavor cues (e.g., almond, banana) were separately mixed in solution with one nutrient (e.g., 10% sucrose) while a second set of flavor cues (e.g., strawberry, vanilla) were separately paired with a second nutrient (e.g., 10% Polycose). Following this training phase, one of the flavor cues paired with each nutrient was then separately extinguished over a number of drinking sessions (e.g., almond, strawberry). Following this, one of the nutrient USs was then devalued by pairing this nutrient with LiCl injections, while the other nutrient US was presented on alternate days without any injection. Thus, one of the nutrient USs was devalued and the value of the other was maintained. Finally, in separate test sessions the animals were confronted with a choice between the two flavor cues that were paired with the same nutrient (i.e., almond versus banana, strawberry versus vanilla). In these tests the animals were given a choice between a non-extinguished and an extinguished flavor cue that were both paired with the same nutrient that had either been devalued or not devalued. The results (see bottom of Figure 3) indicated that animals preferred the flavor cue that was not extinguished when these had been associated with the valued nutrient, but they preferred the extinguished flavor cue to the non-extinguished one when the associated nutrient had been devalued. The results are entirely consistent with and strongly point to the conclusion that the flavor cues evoke specific representations of the nutrients with which they were paired, but that extinction reduces this tendency. If the associated nutrient was devalued then the rats should prefer the flavor cue that “reminds” them less of the nausea-inducing nutrient taste. However, if the associated nutrient was still valuable, then the rats should prefer the flavor cue that more strongly reminds them of the taste of that valuable nutrient.

The results were compelling. Of additional interest, however, was consideration of the role of the motivational state during training and testing. Harris, Gorissen, Bailey, & Westbrook (2000) had earlier provided evidence to suggest that hungry rats are more strongly controlled by flavor-calorie associations while non-hungry rats are more strongly controlled by associations of the other sort described above (it is not clear which from their experimental designs). The pattern of results I described above was observed for rats trained and tested thirsty. Another experiment was performed in that report (Delamater, 2007) with rats that were hungry throughout, and their results were somewhat different. In this case, rats strongly preferred the non-extinguished to the extinguished flavor cue when the associated nutrient was valued. However, when the associated nutrient had been devalued the rats again preferred the nonextinguished to the extinguished flavor cue, but to a lesser degree. In other words, nutrient devaluation in this case diminished the preference for the non-extinguished flavor over the extinguished one, rather than produce a preference for the extinguished flavor.

Delamater (2007) interpreted these results to mean that extinction reduced control not only by the specific flavor-taste association but by the flavor-calorie association as well, and that the latter association more strongly contributed to the preference in hungry rats. To understand this, consider what should have occurred if the rats' preferences were *only* governed by the status of their flavor-taste associations. In this case, the same pattern of results as those reported above for thirsty rats should have been obtained. However, if the rats' preferences were *only* governed by flavor-calorie associations, then the rats should have preferred the non-extinguished to the extinguished flavor cue equally strongly when the associated nutrient was devalued or not. In other words, only an extinction effect and no devaluation effect would be expected. The reason for this is that devaluation reduced the value specifically of the taste of the nutrient, not its caloric significance. However, if both of these associations (flavor-taste, flavor-calorie) contributed equally in hungry rats, then there should have been no preference for either extinguished or non-extinguished flavor cues when the associated nutrient was devalued because each type of association would have led them to prefer a different flavor cue. However, the results indicated a continued preference for the non-extinguished flavor over the extinguished one, albeit to a lesser degree. The reduced preference can only be understood if the flavor-taste associations governed performance to some degree. But the fact that this preference was merely reduced and not reversed suggests that the flavor-calorie associations contributed more strongly than the flavor-taste associations. The fact that rats preferred the non-extinguished to the extinguished flavors implies that extinction weakened control by both of these forms of association.

In a more recent set of studies Delamater (2011) extended these findings to situations in which the flavor cues were nonreinforced in partial reinforcement or latent inhibition procedures. In one study thirsty rats were trained to associate two flavor cues (almond, banana) with the same nutrient (sucrose), but, in addition, one of the flavor cues was presented alone interspersed throughout the conditioning phase (as in a standard partial reinforcement procedure with extra nonreinforced CS presentations). Subsequently, half of the rats received sucrose-LiCl pairings to devalue sucrose whereas the others received sucrose and LiCl unpaired. When rats were given a preference test between the two flavor cues they preferred the consistently reinforced flavor cue when sucrose was still valued but they preferred the partially reinforced flavor cue when sucrose was devalued. The same pattern of results was reported in a second experiment in which all nonreinforced presentations of the flavor cue occurred prior to flavor-nutrient pairings, as in a latent inhibition procedure. In this case, however, more nonreinforced preexposures and fewer flavor-nutrient pairings were required to see the effect. In both cases, however, the effect of nonreinforcement was to diminish control by the specific flavor-taste association during the preference test.

In one further paper we explored the effects of nonreinforced flavor presentations but in the context of a reversal task. Initially, rats learned to associate two different flavors with distinct nutrients (e.g., almond-sucrose, banana-Polycose). Then rats were trained on a reversal of this (e.g., almond-Polycose, banana-sucrose). Notice that during this reversal phase the flavor cues are no longer paired with their original nutrients, and, in this sense, reversal training is like an extinction procedure. We next determined if training on the reversed associations weakened control by the first-learned associations, as in extinction, by devaluing one of the nutrients and determining if rats would display a preference for one flavor cue over the other. In this test, rats avoided the flavor cue that was most recently associated with the devalued nutrient. This result entirely agrees with the findings reported above for extinction, partial reinforcement, and latent inhibition. In all cases, presenting a flavor cue without its associated nutrient weakened control by the specific flavor-taste association. What we have not been able to answer with these tasks, however, is the question of the source of this nonreinforcement effect.

Two possible explanations could be offered for our findings. First, the effects of nonreinforcement (or reversal training) could have partially weakened the specific flavor-taste associations in a manner that would be exactly predicted by the Rescorla-Wagner model. Second, it is possible that extinction produces new learning that masks the expression of the original learning, as seems to also occur in other more conventional learning paradigms (e.g., Bouton, 2004). In both cases, control by the specific flavor-taste associations would be diminished by all of our treatments.

In order to address this issue with our reversal task, we introduced a manipulation that has been shown in other preparations to result in unmasking of an originally learned association following training with an alternative US. In particular, in a subsequent experiment we included a group trained in the same way as I have described above. A second group of rats were given original training and reversal training as above; however, a three-week rest period intervened between reversal training and the selective nutrient devaluation phase. The training of the rats was staggered such that all rats received nutrient devaluation and testing at the same time. We hypothesized that if the originally learned associations were weakened by reversal training that imposing a delay between reversal and testing would be inconsequential. However, if reversal training merely masks expression of those initially learned associations, then control by these associations might recover over time as in spontaneous recovery (see, Bouton & Peck, 1992; Lipatova, Wheeler, Vadillo, & Miller, 2006). The results were consistent with this latter possibility. In particular, we observed that preferences were controlled by the most recently learned associations in the replication group of rats, but rats devalued and tested after a delay displayed preferences that reflected control by the initially learned associations. In other words, these rats avoided the flavor that was originally associated with the devalued nutrient, exactly the opposite pattern of results to those we found in the rats devalued and tested soon after the reversal phase. These findings impressively point to the operation of a masking process of some kind during reversal training, a process that itself is subject to weakening over a delay interval. However, an important caveat is that we cannot yet know whether the same process would be at work in the other procedures we have explored – extinction, partial reinforcement, latent inhibition. It is possible that different processes are recruited in these different tasks. Indeed, we have attempted to provide evidence for spontaneous recovery of control by specific flavor-taste associations following extinction, but have been unable to provide evidence for this thus far. Further work will be required to reach more closure on this point.

Extinction as Conditional Control

Much of the ideas and work described above was motivated by an interest in identifying some of the more basic processes of extinction. As should be clear, it is not always obvious that relatively simple mechanisms (associative weakening, development of inhibitory learning, protection from extinction, summation processes) are absent in many extinction studies. However, as the work of Bouton (e.g., 1991, 2004) has made clear, the story of extinction is more complex than was ever imagined from consideration of the Rescorla-Wagner model alone. In particular, Bouton has suggested that extinction may also entail higher order conditional learning mechanisms. In particular, he suggested that extinction be considered as analogous to a “negative occasion setting” task (e.g., Holland, 1985; Rescorla, 1985). When one admits that the independence of path assumption of the Rescorla-Wagner model is incorrect, then stimuli could concurrently develop associations with the US that are excitatory and inhibitory in nature. From the point of view of predicting behavior, the problem then becomes determining when the excitatory and when the inhibitory association controls performance. Bouton solved the problem by assuming that contextual cues can work like retrieval cues to promote the activation of one or the other associative link. In this sense, the context is said to “modulate” in some way simple conditioning processes.

There has been much effort at detailing what is actually learned in Pavlovian occasion setting tasks (e.g., see Delamater, Kranjec, & Fein, 2010). The ingenious suggestion by Bouton was that extinction, too, involved learning mechanisms analogous to those seen in occasion setting tasks. The clearest example comes from the renewal task described above. In ABA renewal, recall that the CS is trained in one context, extinguished in a second, and then tested either in the extinction or training context. More conditioned responding is seen to the CS when it is tested in its training than its extinction context. As noted above, the Rescorla-Wagner model can, in principle, explain this effect. However, the occasion setting model explains this in a different way. While there are different specific mechanisms proposed to explain occasion setting (Holland, 1985; Rescorla, 1985; Pearce, 1987; 1994; Schmajuk, Lamoureux, & Holland, 1998; Wagner, 2003; Wagner & Brandon, 1989) the most popular account of renewal assumes that the CS develops independent excitatory and inhibitory associations with the US and the extinction context modulates the inhibitory association. Specifically, when the CS is tested in the extinction context, that context is assumed to strengthen the retrievability of the inhibitory CS-US associative link. In contrast, when the CS is tested outside of this context, then the inhibitory CS-US link is not strongly activated, but the more context-general excitatory CS-US link is strongly activated.

While this model differs from the Rescorla-Wagner account of renewal, it is a rather surprising fact that frequently experimental designs are chosen to study renewal that would not allow for a clear separation between these two views. We have recently been exploring renewal phenomena using experimental designs that more clearly point to the involvement of an occasion setting like process. In one set of studies, Delamater, Campese, & Westbrook (2009) examined ABA renewal and compared this to an ABB control condition, but we chose an experimental design that controls for the separate conditioning histories of the experimental contexts (see Figure 4). In this way, it would not be easy for the sorts of simple mechanisms identified by the Rescorla-Wagner model to apply. In particular, we trained rats in an appetitive magazine approach task to associate one stimulus with food pellets in one experimental context and a second stimulus with food pellets in a second experimental context (i.e., Ctx 1: CS1-US, Ctx 2: CS2-US). During the extinction phase, each CS was extinguished in the other CS’s acquisition context (i.e., Ctx 1: CS2−, Ctx 2: CS1−). Finally, each CS was tested in both contexts to assess renewal (i.e., Ctx 1: CS1− & CS2−, Ctx 2: CS1− & CS2−). There are a couple of noteworthy features of this design. First, notice that each CS is tested in its training context (ABA renewal) and also in its extinction context (ABB control). Thus, renewal is assessed to each stimulus using a sensitive within-subject procedure. Second, each context plays the same role as an acquisition context for one stimulus and an extinction context for another stimulus. Thus, the associative histories of each context are equivalent and, as a result, any direct associations that the contexts may have with the pellet US (be they excitatory or inhibitory) should be comparable. Because of this fact, it becomes difficult for the Rescorla-Wagner model to explain differences in responding to the CSs when they are tested in ABA and ABB conditions. A conceptually similar experimental design was used by Lovibond, Preston, & Mackintosh (1984) and they failed to observe renewal. This result prompted those authors to speculate that it may be unwise to completely dismiss the possibility that Rescorla-Wagner mechanisms may contribute to renewal phenomena. In our task, however, we observed significantly greater magazine approach responding when the CSs were tested under ABA than ABB conditions. Our interpretation was that this experimental design more clearly illustrates the operation of occasion-setting mechanisms in renewal.

In a recent set of studies (unpublished) a former PhD student of mine, Vincent Campese, used this general approach in his doctoral dissertation research to examine the role of the dorsal hippocampus in controlling renewal phenomena. Based on similar work conducted in the conditioned freezing paradigm (Corcoran & Maren, 2001; 2004; Maren & Hobin, 2007), we speculated that the dorsal hippocampus might be especially important in ABC but perhaps less so in ABA renewal. We used the same experimental design just described, but prior to testing the rats received infusions of muscimol (a GABAa receptor agonist that causes widespread cellular inactivation) into the dorsal hippocampus. Control rats received saline infusions during these tests. We observed that this manipulation had no impact on ABA renewal, although the manipulation did impair subjects' performance on a spatial task that has been shown in prior research to be sensitive to dorsal hippocampal inactivation.

In a further experiment, we examined the role of the dorsal hippocampus in ABC renewal. In this task, two stimuli were initially trained in one context (Ctx 1: CS1-US, CS2-US) before each CS was extinguished in different contexts (Ctx 2: CS1−, Ctx 3: CS2−). Then, ABC renewal was assessed by testing each CS in each of the two extinction contexts (Ctx 2: CS1− & CS2−, Ctx 3: CS1− & CS2−). Note that this design, as well, controls for the associative histories of the two extinction contexts, making it difficult for any mechanism other than occasion setting to account for renewal in this case. We observed greater responding to the CSs when they were tested in the context in which a different CS was extinguished than when tested in the context in which they were extinguished. Furthermore, once again, dorsal hippocampal inactivation had no impact on this renewal effect (although they did impair, once again, spatial performance in these rats). Whereas earlier fear conditioning studies showed that dorsal hippocampal inactivation impaired ABC renewal and dorsal hippocampal lesions impaired ABA renewal, it is important to note that those experiments did not always use experimental designs that more specifically target occasion setting processes. Thus, it is not always possible when one obtains evidence for lesion or inactivation effects to conclude that the locus of those effects is on the occasion setting process, per se. There could, of course, very well be a difference in the neural mechanisms of appetitive and aversive renewal and this would account for our divergent sets of results. At the same time, however, some of the discrepancies might also be accounted for in terms of the particular processes that are likely to play a role in the different experimental designs. This last point is a very important one to keep in mind when attempting to identify through various brain manipulations the functional significance of different brain structures.

Closing Thoughts and Conclusions

The Rescorla-Wagner model (1972) of Pavlovian learning has enjoyed a tremendous amount of success in helping learning theorists understand a wide variety of phenomena. The model has also served to stimulate theory development by serving as a point of contrast in evaluating more modern approaches. While a large number of learning theories have appeared in the time since this theory was proposed, current investigations into the neural mechanisms of learning have provided convincing evidence that supports some of the Rescorla-Wagner model’s most basic assumptions. Chief among these include the “prediction error” concept and its role in driving learning. Studies have revealed that individual neurons possess the property of responding when important events are unpredicted but not otherwise. Furthermore, recent work examining the neural circuits of learning is revealing mechanisms that function to limit processing of the US in ways anticipated by the Rescorla-Wagner model.

In spite of these supportive findings, there remain a number of critical challenges to some of the basic premises of the Rescorla-Wagner model. Perhaps the most significant challenges come from research on Pavlovian extinction. While the model anticipates that extinction should result in some unlearning, it is important to realize that under many circumstances the model does not predict that extinction should result in complete unlearning. Thus, phenomena that point to the persistence of learning after extinction (e.g., spontaneous recovery, renewal, reinstatement) should not, in and of themselves, be taken as evidence against the model’s account of extinction. However, other research has shown that under many circumstances extinction may not even result in any weakening of excitatory associative strength (Delamater, 1996; Rescorla, 1996).

In some of our more recent work we have begun to find evidence to suggest that control by sensory-specific CS-US associations is undermined by extinction. In particular, when extinction occurs following a more limited amount of training the extinction treatment may have a more enduring effect. It is tempting to hypothesize that extinction will be more effective when a CS-US association has been less well encoded from the outset (see also Eisenberg, et al, 2003). Two important issues concerning this possibility include (1) determining if the loss of control by sensory-specific CS-US associations reflects a true weakening of those associations or masking, and (2) determining if other procedures that result in poor encoding might also be more sensitive to extinction treatments. Clearly, more research will be required to address these issues.

Another important issue suggested by our work on extinction concerns the possibility that multiple forms of associative learning might be affected in different ways by extinction. It is now quite common to understand Pavlovian learning in terms of associations between the CS and multiple components of the US, e.g., its sensory, emotional, and specific response components (e.g., Delamater & Oakeshott, 2007; Konorski, 1967; Wagner & Brandon, 1989). There has been very little attention in basic extinction research directed to the question of whether these different aspects of learning might obey different learning rules. To the extent that these different aspects of learning represent truly distinct learning processes, it would not at all be surprising if they would each show different sensitivities to extinction treatments. If this were true, then the translation of basic extinction research into therapeutic practice would require an understanding of what response systems would and would not be expected to be affected by extinction treatments. More research will be necessary to address this possibility. My earlier finding that conditioned magazine approach responses show a more enduring effect of extinction whereas sensory-specific PIT effects appear unaffected could indicate nothing other than a sensitivity difference in the different response systems to extinction. On the other hand, these findings could reflect, instead, true differences in the way in which extinction influences different forms of learning.

Our research has also been motivated by an interest in comparing extinction effects in different learning paradigms – magazine approach and flavor preference learning. The flavor preference learning paradigm is particularly interesting to examine because of the fact that learned preferences are often reported to be extremely resistant to extinction (e.g., Drucker, et al, 1994; Harris, et al., 2004). This is highly unusual because even in more standard learning paradigms extinction is routinely observed to diminish conditioned responding. If learned preferences are truly unaffected by extinction treatments this could support the interpretation that this form of learning is unique (e.g., De Houwer, et al, 2001; Pearce, 2002). Our own research on this topic, however, has shown that flavor preference learning appears quite sensitive to extinction (and other treatments involving nonreinforcement) when one assesses such effects by comparing extinguished to non-extinguished cues. The more typical way of assessing extinction in this paradigm involves determining if a learned preference for a flavor CS+ over either another flavor CS− or plain water is reduced by the treatment. This method may be especially insensitive because once a preference has been established any residual learning about that flavor cue may be sufficient to result in a preference for that flavor relative to a more neutral alternative. On the other hand, by directly comparing preference for a flavor that has or has not undergone extinction, any relative difference in associative values should be more easily detected. Unfortunately, this approach is almost never used, and, rather different conclusions may be drawn about the nature of the learning process as a result.

Another important set of questions concerning extinction in flavor preference learning has to due with the issue, noted above, that extinction may work differently on different components of learning. The flavor preference paradigm is an interesting one in this regard because flavor cues are thought to associate with several distinct features of the nutrient US, e.g., its taste, hedonic, and/or motivational components. Here, once again, there is the opportunity to assess the effects of extinction on distinct aspects of learning. More research will be required, though, in order to provide better answers to these questions than we currently have.

A fundamentally important problem in extinction research, regardless of what paradigm we use, continues to be whether extinction results in true associative weakening or masking. It is not clear how this sort of issue will ever be resolved, however. Strong proponents of the weakening view will be able to maintain that some weakening occurs even in situations where spontaneous recovery, renewal, and reinstatement effects can be found. These phenomena are often cited as providing evidence against unlearning and for a masking view. It should be obvious, however, that a conventional approach (like the Rescorla-Wagner model) could very well anticipate these phenomena. However, the total absence of these phenomena would appear to be more consistent with an associative weakening (or unlearning) view (see Kim and Richardson, 2007a; 2007b; Monfils, Cowansage, Klann, & LeDoux, 2009; Myers, Ressler, & Davis, 2006). It should be noted that even here, a strong proponent of the masking view could still maintain that the manipulation simply results in more effective masking – the treatment, in other words, makes it especially easy to retrieve the memory of extinction. This being the case, it seems like convincing evidence to settle this dispute would be hard-pressed to obtain. Perhaps, a better approach would be to develop specific theoretical models that could be tested under conditions where different models lead to opposing predictions. One promising approach to this was suggested by Mauk and his colleagues in their attempt to model extinction in the eyeblink conditioning neural circuit. In this model, extinction is assumed to result in weakening of underlying learned connections at multiple loci within the neural network, but global characteristics of the network reflect retention of associative information even well after complete response loss has occurred (Mauk & Ohyama, 2004; also Kehoe, 1988; Larrauri & Schmajuk, 2008). Thus, extinction very likely reflects the operation of both some “unlearning” process as well as retention of learning. More targeted tests of models of this general sort could be very beneficial in future extinction research.

Finally, one of the most exciting and positive developments arising from early challenges to the Rescorla-Wagner approach to extinction has been the suggestion that conditional control mechanisms (like occasion setting) apply to extinction. Bouton and his colleagues have demonstrated that contexts can act as occasion setters in modulating performance to extinguished CSs (e.g., Bouton, 2004; see also Harris, Jones, Bailey, & Westbrook, 2000). This work is compelling and strongly points to weaknesses in approaches that only emphasize associative strengthening and weakening mechanisms. Instead, extinction can be profitably viewed as involving additional complex “hierarchical” (occasion setting) associative mechanisms. However, one important point to notice in connection with this is that just because occasion-setting processes may be engaged by extinction, this does not imply that simpler mechanisms also do not apply in any given situation. Indeed, our choice of experimental design is crucial here because if one uses an experimental design that could equally recruit either type of mechanism, then either type of mechanism could explain the results. The issue become of paramount importance when we are looking to establish the functional significance of different brain regions.

In our own work looking at the role of the dorsal hippocampus in renewal we have found results that are at odds with conclusions based on fear conditioning studies. There are many ways in which appetitive and aversive learning situations differ, but the general point here is that it is also possible that different results could reflect different psychological mechanisms engaged by different tasks. We will need to be very careful in using tasks that allow us to hone in on the particular psychological mechanism that we are targeting. It is sometimes taken for granted that since extinction is thought to involve occasion setting mechanisms, all extinction experimental designs must reflect control by occasion setting processes. This may be a mistake and could lead to erroneous conclusions about the functional significance of particular brain structures.

In summary, I have revisited the Rescorla-Wagner model account of extinction and suggested that its most basic assumption, that prediction errors drive learning, receives a fair amount of support from neuroscience analyses. At the same time, I have pointed out that data that have traditionally been taken as evidence against the Rescorla-Wagner view of extinction (spontaneous recovery, renewal, reinstatement) are not necessarily at odds with such an account. However, the view faces difficulty, in particular, with results showing that extinction does not appear to weaken specific associations at all. Additional work will be needed to more fully understand when extinction treatments do and do not undermine control by sensory-specific associations. Encoding level and response system variables may be critical here. Our work has also shown procedures in which extinction effects can readily be obtained in flavor preference learning, a result that is consistent with other more traditional learning paradigms. Finally, we have also presented additional data that strongly point to the operation of occasion setting like processes in extinction. Overall, our work suggests that the choice of paradigm and experimental design are both extremely important given the complexity of the types of processes that can be engaged by simple extinction procedures.

Acknowledgments

This research reported here was supported by National Institute of Mental Health (065947) and Professional Staff Congress – City University of New York (62413-00 40, 63606-00 41) grants awarded to the author.

References

Berridge KC, Grill HJ. Isohedonic tastes support a two-dimensional hypothesis of palatability. Appetite. 1984; 5:221–231. [PubMed: 6524918]

Blundell P, Hall G, Killcross S. Preserved sensitivity to outcome value after lesions of the basolateral amygdala. Journal of Neuroscience. 2003; 23:7702–7709. [PubMed: 12930810]

Bouton, ME. Context and retrieval in extinction and in other examples of interference in simple associative learning. In: Dachowski, L.; Flaherty, CF., editors. Current topics in animal learning: Brain, emotion, and cognition. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc; 1991. p. 25-53.

Bouton ME. Context and behavioral processes in extinction. Learn Mem. 2004; 11:485–494. [PubMed: 15466298]

Bouton ME, King DA. Contextual control of the extinction of conditioned fear: tests for the associative value of the context. J Exp Psychol Anim Behav Process. 1983; 9(3):248–265. [PubMed: 6886630]

Bouton ME, King DA. Effect of context on performance to conditioned stimuli with mixed histories of reinforcement and nonreinforcement. Journal of Experimental Psychology: Animal Behavior Processes. 1986; 12(1):4–15.

Bouton ME, Peck CA. Spontaneous recovery in cross-motivational transfer (counter conditioning). Animal Learning & Behavior. 1992; 20:313–321.

Bouton ME, Swartzentruber D. Analysis of the associative and occasion-setting properties of contexts participating in a Pavlovian discrimination. Journal of Experimental Psychology: Animal Behavior Processes. 1986; 12(4):333–350.

Bouton ME, Westbrook RF, Corcoran KA, Maren S. Contextual and Temporal Modulation of Extinction: Behavioral and Biological Mechanisms. Biological Psychiatry. 2006; 60:352–360. [PubMed: 16616731]

Brandon SE, Wagner AR. Modulation of a discrete Pavlovian conditioned reflex by a putative emotive Pavlovian conditioned stimulus. Journal of Experimental Psychology: Animal Behavior Processes. 1991; 17:299–311. [PubMed: 1890388]

Corcoran KA, Maren S. Hippocampal inactivation disrupts contextual retrieval of fear memory after extinction. J Neurosci. 2001; 21:1720–1726. [PubMed: 11222661]

Corcoran KA, Maren S. Factors regulating the effects of hippocampal inactivation on renewal of conditional fear after extinction. Learning & Memory. 2004; 11:598–603. [PubMed: 15466314]

Delamater AR. Outcome-selective effects of intertrial reinforcement in a Pavlovian appetitive conditioning paradigm with rats. Animal Learning & Behavior. 1995; 23(1):31–39.

Delamater AR. Effects of several extinction treatments upon the integrity of Pavlovian stimulus-outcome associations. Animal Learning & Behavior. 1996; 24:437–449.

Delamater AR. Experimental Extinction in Pavlovian Conditioning: Behavioural and Neuroscience Perspectives. The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology. 2004; 57:97–132.

Delamater AR. Extinction of Conditioned Flavor Preferences. Journal of Experimental Psychology: Animal Behavior Processes. 2007; 33:160–171. [PubMed: 17469964]

Delamater AR. On the nature of CS and US representations in Pavlovian learning. Learning & Behavior. 2012; 40:1–23. [PubMed: 21786019]

Delamater AR, Campese V, LoLordo VM, Sclafani A. Unconditioned Stimulus Devaluation Effects in Nutrient-Conditioned Flavor Preferences. Journal of Experimental Psychology: Animal Behavior Processes. 2006; 32:295–306. [PubMed: 16834496]

Delamater AR, Campese V, Westbrook RF. Renewal and spontaneous recovery, but not latent inhibition, are mediated by gamma-aminobutyric acid in appetitive conditioning. Journal of

Experimental Psychology: Animal Behavior Processes. 2009; 35(2):224–237. [PubMed: 19364231]

Delamater AR, Holland PC. The influence of CS-US interval on several different indices of learning in appetitive conditioning. Journal of Experimental Psychology: Animal Behavior Processes. 2008; 34:202–222. [PubMed: 18426304]

Delamater AR, Kranjec A, Fein M. Differential outcome effects in Pavlovian biconditional and ambiguous occasion setting tasks. Journal of Experimental Psychology: Animal Behavior Processes. 2010; 36:471–481. [PubMed: 20718549]

Delamater AR, Oakeshott S. Learning about multiple attributes of reward. In B. Balleine, K Doya, J. O'Doherty, & M. Sakagami (Eds.) Reward and Decision Making in Cortico-basal Ganglia Networks. Annals of the New York Academy of Sciences. 2007; 1104:1–20. [PubMed: 17344542]

Drucker DB, Ackroff K, Sclafani A. Nutrient-conditioned flavor preference and acceptance in rats: effects of deprivation state and nonreinforcement. Physiology and Behavior. 1994; 56:701–707. [PubMed: 7800736]

Dwyer DM. Reinforcer devaluation in palatability-based learned flavor preferences. Journal of Experimental Psychology: Animal Behavior Processes. 2005; 31:487–492. [PubMed: 16248735]

Eisenberg M, Kobilo T, Berman DE, Dudai Y. Stability of retrieved memory: Inerse correlation with trace dominance. Science. 2003; 301:1102–1104. [PubMed: 12934010]

Elizalde G, Sclafani A. Flavor Preferences Conditioned by Intragastric Polycose Infusions: A Detailed Analysis Using an Electronic Esophagus Preparation. Physiology & Behavior. 1990; 47:63–77. [PubMed: 2109327]

Gallistel CR, Gibbon J. Time, rate, and conditioning. Psychological Review. 2000; 107(2):289–344. [PubMed: 10789198]

Grau JW, Rescorla RA. Role of context in autoshaping. Journal of Experimental Psychology: Animal Behavior Processes. 1984; 10:324–332.

Harris JA, Gorissen MC, Bailey GK, Westbrook RF. Motivational state regulates the content of learned flavor preferences. Journal of Experimental Psychology: Animal Behavior Processes. 2000; 26:15–30. [PubMed: 10650541]

Harris JA, Jones ML, Bailey GK, Westbrook RF. Contextual control over conditioned responding in an extinction paradigm. J Exp Psychol Anim Behav Process. 2000; 26(2):174–185. [PubMed: 10782432]

Harris JA, Shand FL, Carroll LQ, Westbrook RF. Persistence of preference for a flavor presented in simultaneous compound with sucrose. Journal of Experimental Psychology: Animal Behavior Processes. 2004; 30:177–189. [PubMed: 15279509]

Holland, PC. The nature of conditioned inhibition in serial and simultaneous feature negative discriminations. In: Miller, RR.; Spear, NE., editors. Information processing in animals: Conditioned inhibition. Hillsdale NJ: Lawrence Erlbaum Associates Inc; 1985. p. 267-298.

Kehoe EJ. A layered network model of associative learning: Learning to learn and configuration. Psychological Review. 1988; 95(4):411–433. [PubMed: 3057526]

Kim JH, Richardson R. A developmental dissociation of context and GABA effects on extinguished fear in rats. Behavioral Neuroscience. 2007a; 121:131–139. [PubMed: 17324057]

Kim JH, Richardson R. A developmental dissociation in reinstatement of an extinguished fear in rats. Neurobiology of Learning & Memory. 2007b; 88:48–57. [PubMed: 17459734]

Kim JJ, Krupa DJ, Thompson RF. Inhibitory cerebello-olivary projections and blocking effect in classical conditioning. Science. 1998; 279(5350):570–573. [PubMed: 9438852]

Konorski, J. Integrative activity of the brain. Chicago: University of Chicago Press; 1967.

Kruse JM, Overmier JB, Konz WA, Rokke E. Pavlovian conditioned stimulus effects upon instrumental choice behavior are reinforcer specific. Learning & Motivation. 1983; 14(2):165–181.

Larrauri JA, Schmajuk NA. Attentional, associative, and configural mechanisms in extinction. Psychological Review. 2008; 115:640– 676. [PubMed: 18729595]

Lipatova O, Wheeler DS, Vadillo MA, Miller RR. Recency to primacy shift in cue competition. Journal of Experimental Psychology: Animal Behavior Processes. 2006; 32:396–406. [PubMed: 17044742]

Lovibond PF, Preston GC, Mackintosh NJ. Context Specificity of Conditioning, Extinction, and Latent Inhibition. Journal of Experimental Psychology: Animal Behavior Processes. 1984; 10:360–375.

Maren S, Hobin JA. Hippocampal regulation of context-dependent neuronal activity in the lateral amygdala. Learning & Memory. 2007; 14:318–324. [PubMed: 17522021]

Maren S, Quirk GJ. Neuronal signaling of fear memory. Nature Reviews Neuroscience. 2004; 5:844–852.

Mauk MD, Ohyama T. Extinction as new learning versus unlearning: Considerations from a computer simulation of the cerebellum. Learning & Memory. 2004; 11:566–571. [PubMed: 15466310]

McNally GP, Johansen JP, Blair HT. Placing prediction into the fear circuit. Trends in Neurosciences. 2011; 34:283–292. [PubMed: 21549434]

Miller RR, Barnet RC, Grahame NJ. Assessment of the Rescorla-Wagner model. Psychological Bulletin. 1995; 117:363–386. [PubMed: 7777644]

Monfils MH, Cowansage KK, Klann E, LeDoux JE. Extinction-reconsolidation boundaries: Key to persistent attenuation of fear memories. Science. 2009; 324:951–955. [PubMed: 19342552]

Myers KM, Davis M. Mechanisms of fear extinction. Molecular Psychiatry. 2007; 12:120–150. [PubMed: 17160066]

Myers KM, Ressler KJ, Davis M. Different mechanisms of fear extinction dependent on length of time since fear acquisition. Learning & Memory. 2006; 13:216–223. [PubMed: 16585797]

Nader K, Schafe GE, Le Doux JE. Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. Nature. 2000; 406(6797):722–726. [PubMed: 10963596]

Pearce JM. A model for stimulus generalization in Pavlovian conditioning. Psychological Review. 1987; 94:61–75. [PubMed: 3823305]

Pearce JM. Similarity and discrimination: A selective review and a connectionist model. Psychological Review. 1994; 101:587–607. [PubMed: 7984708]

Pearce JM. Evaluation and development of a connectionist theory of configural learning. Anim Learn Behav. 2002; 30(2):73–95. [PubMed: 12141138]

Pearce JM, Hall G. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review. 1980; 87(6):532–552. [PubMed: 7443916]

Polack CW, Laborda MA, Miller RR. Extinction context as a conditioned inhibitor. Learning & Behavior. 2011

Quirk GJ, Mueller D. Neural Mechanisms of Extinction Learning and Retrieval. Neuropsychopharmachology. 2008; 33:56–72.

Rescorla RA. Preservation of Pavlovian associations through extinction. Quarterly Journal of Experimental Psychology: Comparative & Physiological Psychology. 1996; 49:245–258.

Rescorla, RA. Conditioned inhibition and facilitation. In: Miller, RR.; Spear, NE., editors. Information processing in animals: Conditioned inhibition. Hillsdale, NJ: Erlbaum; 1985. p. 299-326.

Rescorla RA. Retraining of extinguished Pavlovian stimuli. J Exp Psychol Anim Behav Process. 2001; 27(2):115–124. [PubMed: 11296487]

Rescorla RA. Protection from extinction. Learn Behav. 2003; 31(2):124–132. [PubMed: 12882371]

Rescorla, RA.; Wagner, AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black, AHP.; Prokasy, WF., editors. Classical conditioning II: Current research and theory. New York: Appleton-Century-Crofts; 1972. p. 64-99.

Rodger RS. Multiple contrasts, factors, error rate, and power. British Journal of Mathematical & Statistical Psychology. 1974; 27:179–198.

Schmajuk NA, Lamoureux JA, Holland PC. Occasion setting: A neural network approach. Psychological Review. 1998; 105(1):3–32. [PubMed: 9450370]

Schoenbaum G, Roesch MR, Stalnaker TA, Takahashi YK. A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. Nature Reviews Neuroscience. 2009; 10:885–892.

Sotres-Bayon F, Bush DE, LeDoux JE. Emotional perseveration: an update on prefrontal-amygdala interactions in fear extinction. Learn Mem. 2004; 11(5):525–535. [PubMed: 15466303]

Spence KW, Rutledge EF, Talbott JH. Effect of number of acquisition trials and the presence or absence of the UCS on extinction of the eyelid CR. Journal of Experimental Psychology. 1963; 66:286–291. [PubMed: 14054385]

Swartzentruber D, Rescorla RA. Modulation of trained and extinguished stimuli by facilitators and inhibitors. Animal Learning & Behavior. 1994; 22(3):309–316.

Tobler PN, Dickinson A, Schultz W. Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. The Journal of Neuroscience. 2003; 23:10402–10410. [PubMed: 14614099]

Waelti P, Dickinson A, Schultz W. Dopamine responses comply with basic assumptions of formal learning theory. Nature. 2001; 412(6842):43–48. [PubMed: 11452299]

Wagner AR. Context-sensitive elemental theory. Quarterly Journal of Experimental Psychology. 2003; 23B:7–29. [PubMed: 12623534]

Wagner, AR.; Brandon, SE. Evolution of a structured connectionist model of Pavlovian conditioning (AESOP). In: Klein, SB.; Mowrer, RR., editors. Contemporary learning theories: Pavlovian conditioning and the status of traditional learning theory. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1989. p. 149-189.

Highlights

- Pavlovian extinction learning: associative weakening or masking?
- Selective Pavlovian-to-instrumental transfer (PIT) tests reveal control by sensory-specific associations
- Control by sensory-specific associations are sometimes sensitive to extinction
- PIT tasks can be used to demonstrate effects of extinction
- US devaluation tasks also can be used to show that control by sensory-specific associations are sensitive to extinction
- Extinction can also be shown to be influenced by conditional control (occasion setting) processes

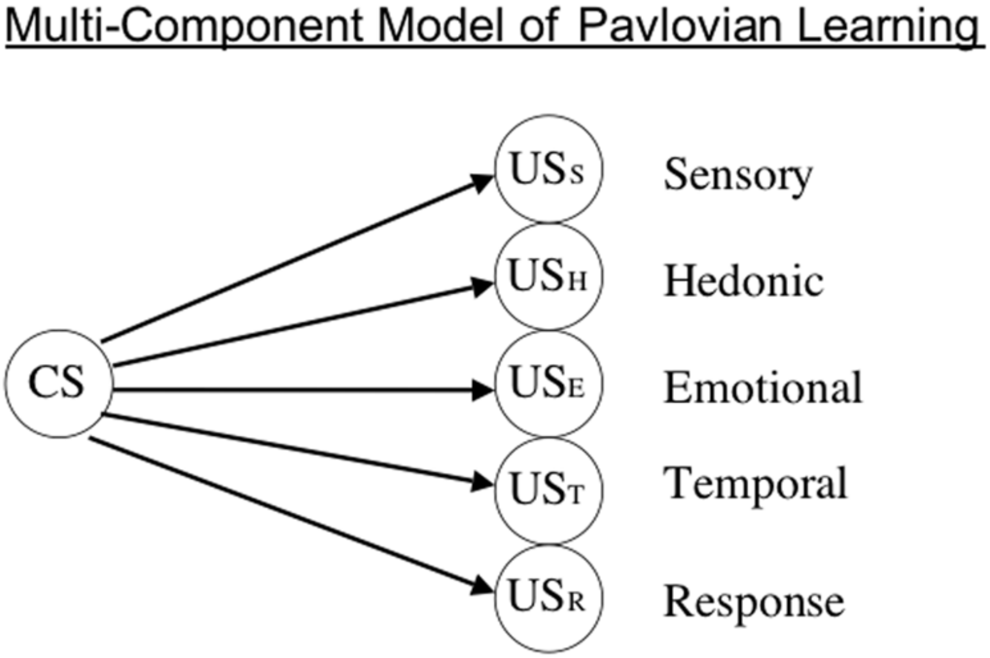


Figure 1.

The multi-component model of Pavlovian learning shows that representations of the CS can enter into independent associations with separate sensory, hedonic, emotional, temporal, and response components of the US.

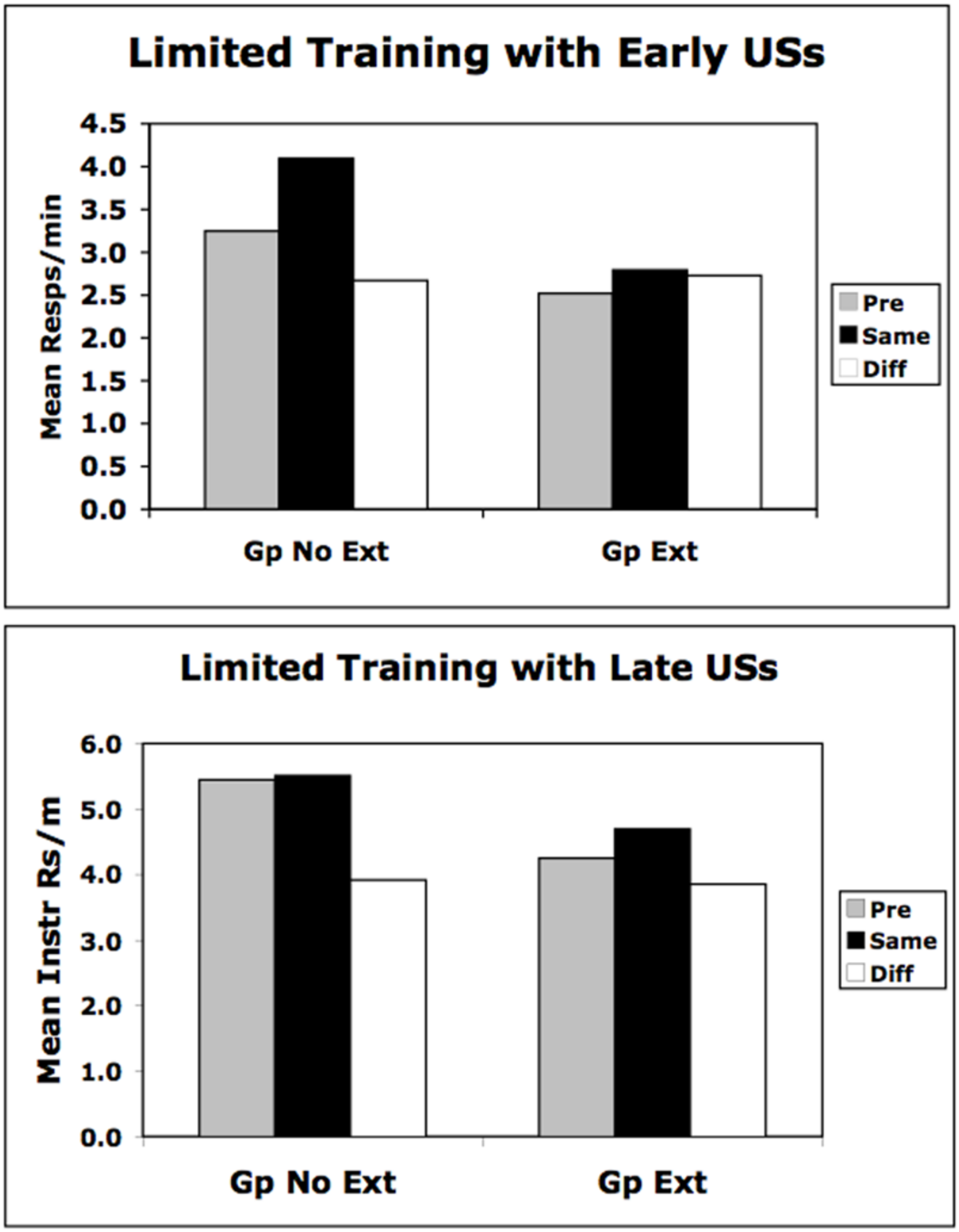


Figure 2.

Data from Pavlovian-to-instrumental transfer tests showing instrumental responding during the Pre stimulus period and also during the stimulus but separated in terms of the response that was previously reinforced with the same or different outcome as that signaled by the CS. The data are shown for separate groups of rats given a small amount of Pavlovian training and either extensive extinction (Gp Ext) or just context exposure (Gp No Ext). Further, in separate experiments rats were either trained with the US occurring towards the beginning of the 120 s CSs (top) or at the offset of 60 s CSs (bottom). See text for additional details.

Acquisition	Extinction	US Devaluation	Test
A + Nutr 1	A-		
B + Nutr 1		Nutr 1 - LiCl	A vs B
C + Nutr 2	C-	Nutr 2 -	C vs D
D + Nutr 2			

Nutr 1, 2 = 8% Sucrose or 8% Polycose
A - D: Alm, Ban, Van, Str

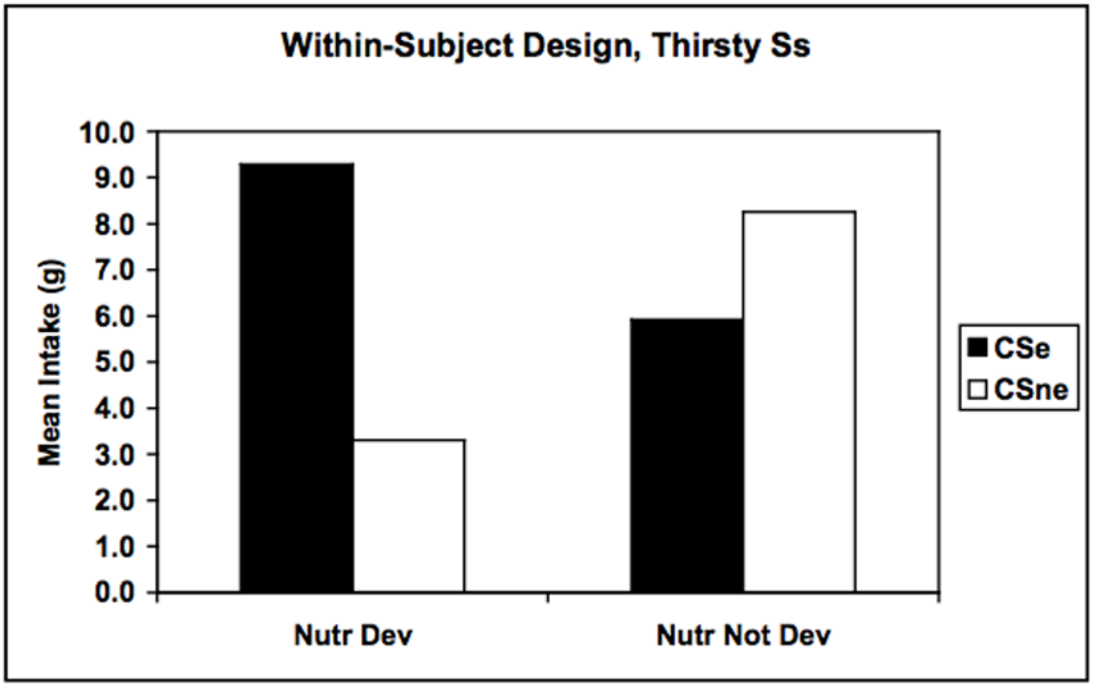
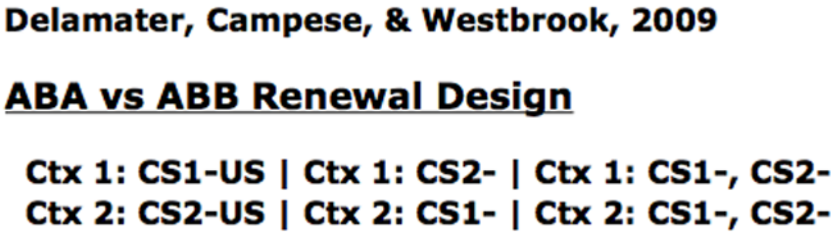


Figure 3.
Experimental design (top) used by Delamater (2007) to study extinction of specific flavor-taste associations in a flavor preference learning paradigm. Each of four distinct flavor cues (A, B, C, D) were separately paired with different nutrients (Nutr 1, Nutr 2) during the Acquisition phase, but two of these was presented without any nutrients during the Extinction phase. One of the nutrients was then devalued by being paired with lithium chloride (LiCl) during the US devaluation phase, and, finally, the rats were given separate preference tests between A and B or C and D. Data from the preference tests appear below. See text for additional details.



Delamater, Campese, & Westbrook, 2009

ABA vs ABB Renewal Design

Ctx 1: CS1-US | Ctx 1: CS2- | Ctx 1: CS1-, CS2-
Ctx 2: CS2-US | Ctx 2: CS1- | Ctx 2: CS1-, CS2-

Figure 4.
Experimental design used by Delamater, Campese, & Westbrook (2009). See text for details.