

Conditioning, Remembering, and Forgetting

Mark E. Bouton

Theories of learning and of memory have been artificially separated for many years. In this article it is argued that the separation should be abandoned. Remembering and forgetting play an important role in simple conditioning experiments. For example, conditioned performance can readily recover after extinction because extinction is susceptible to forgetting. Memory itself involves the kind of associative learning that conditioning experiments are designed to investigate (e.g., conditioning experiments provide insight into the mechanisms of memory retrieval). Learning, remembering, and forgetting all occur within the same biological context; their adaptive functions are therefore intertwined. Taken together, they shed light on some of the mechanisms of clinical relapse. Conditioning theory enriches, and is enriched by, the field of memory as well as other fields of behavioral science.

An organism's ability to learn implies an ability to remember. That is because any demonstration of learning requires a memory in some sense of an experience from the past. The relationship between learning and memory has often been neglected, however; learning theorists, for example, have not always said much about it. Earlier in the history of the field, when learning was viewed as the acquisition of an association between a stimulus and a response, any notion of memory beyond the stimulus–response association itself seemed unnecessary to explain behavior. Conversely, human memory research often has had little to say about learning, focusing instead on what is represented in memory rather than how the memory was acquired in the first place. Current work in animal cognition, with its emphasis on how animals represent things such as space, time, number, and the delayed-matching-to-sample task, likewise seems to ignore acquisition processes. And although there has always been an interest in long-term memory in animals as it is manifest in associative learning tasks (e.g., Spear, 1978), there has been little impact of this interest on what can be called mainstream conditioning theory.

In this article I argue that a blending of ideas about learning and remembering is extremely valuable. Learning theory has much to gain by acknowledging the memory aspects of conditioning experiments. At the same time, memory theory may likewise benefit by recognizing that conditioning, or rather the associative learning mechanisms that conditioning experiments elucidate, plays an equally fundamental role in remembering. I show in this article that

remembering and forgetting have an impact on even the most basic of conditioning phenomena, that memory itself involves the kind of associative learning that conditioning experiments are designed to investigate, and that acknowledging the relationships among conditioning, remembering, and forgetting is useful when one examines animal and human behavior in their natural contexts.

Extinction, Interference, and Memory Retrieval

It is not necessary to look any further than extinction, one of the most fundamental of all behavioral phenomena, to find a role for remembering and forgetting in conditioning. In extinction, a stimulus (a conditioned stimulus [CS]) that signals a biologically significant event (an unconditioned stimulus [US]) loses its ability to evoke responding when it is repeatedly presented alone. Pavlov (1927) recognized that extinction is probably as important as conditioning itself as a way for the animal to adapt to a continually changing environment. Behavior must be able to increase and decrease according to changing contingencies.

One of the simplest explanations of extinction is to assume that presenting the CS without the US merely reduces or destroys what was originally learned (e.g., Rescorla & Wagner, 1972). Pavlov (1927) recognized, however, that extinction does not depend on unlearning. He discovered effects such as spontaneous recovery, in which the extinguished response recovers after extinction if time elapses before the CS is tested again. The fact that the original response can recover after extinction indicates that the original learning could not have been destroyed. Spontaneous recovery is so well known that introductory psychology students often learn about it at the same time they learn about extinction. It is therefore ironic that modern models of conditioning (e.g., Mackintosh, 1975; Pearce, 1987; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Wagner, 1978, 1981) do not account for it or predict it. The effects of retention intervals are the subject of memory theory, not learning theory.

There is other evidence, of course, that extinction is not unlearning. Over the past decade, research in my laboratory

Editor's Note. This article is another in a series on various topics to be invited by the Editor.

Preparation of this article was supported by Grant IBN 92-09454 from the National Science Foundation. I am grateful to Andy Baker, Ralph Miller, Tom Nelson, Michael Speed, Tony Wright, Cody Brooks, Byron Nelson, and Sean Ricker for their comments.

Correspondence concerning this article should be addressed to Mark E. Bouton, Department of Psychology, University of Vermont, Burlington, Vermont 05405.

and in others has shown that performance after extinction is crucially determined by the *context*. In experiments run in my laboratory, context was usually operationally defined as background cues provided by the apparatus in which the CS was presented. As one illustration of how the context affects extinguished performance, consider a phenomenon known as *reinstatement*. After extinction has occurred, one can condition the context by presenting the US in it alone a few times. The resulting contextual conditioning then reinstates extinguished responding to the CS: When the CS is presented in the context, the animal responds to it again (e.g., Bouton, 1984; Bouton & Bolles, 1979b). Interestingly, performance to CSs that have not been extinguished are much less affected by contextual conditioning (e.g., Bouton, 1984; Bouton, Rosengard, Achenbach, Peck, & Brooks, 1993). Performance to an extinguished CS is unusually sensitive to its background. Contextual conditioning may remind the animal of conditioning after extinction has occurred (e.g., Bouton et al., 1993).

Another effect of context on extinction performance is called the *renewal effect*. If conditioning occurs in one context (Context A) and then extinction occurs in another context (Context B), the loss of responding that occurs in extinction happens at a rate that is identical to that observed in a subject for whom the context is not switched. However, responding to the CS returns (is renewed) when the CS is returned to the original conditioning context (e.g., Bouton & Bolles, 1979a; Bouton & King, 1983; Bouton & Peck, 1989; Grahame, Hallam, Geier, & Miller, 1990; Lovibond, Preston, & Mackintosh, 1984). Such renewal occurs after extinction training is extended long past the point at which responding has been eliminated (e.g., Bouton & Swartzentruber, 1989). The phenomenon also occurs when conditioning, extinction, and testing occur in Contexts A, B, and C (e.g., Bouton & Bolles, 1979a; Bouton & Brooks, 1993) and, most recently, in Contexts A, A, and B (Bouton & Ricker, in press). Extinction performance depends on the subject being in the extinction context.

The results of experiments on the renewal effect suggest that the context does not influence performance to the CS through its direct association with the US. It would be easy to explain renewal if the conditioning context were merely associated with the US or if the extinction context acquired direct inhibitory associations with it. Such possibilities are consistent with mainstream models of conditioning (e.g., Rescorla & Wagner, 1972). However, my students and I have found no evidence that the contexts control performance in this sort of paradigm by entering into direct, binary associations with the US (e.g., Bouton & King, 1983; Bouton & Swartzentruber, 1986). Apparently, contextual cues control performance through some mechanism other than a direct context-US association. I have suggested that they retrieve the current "meaning" of the CS (e.g., Bouton, 1991, 1993; Bouton & Bolles, 1985). Extinction does not destroy the original CS-US association. Instead, the CS acquires a second meaning (e.g., a CS-no US association). At the end of extinction, the CS has two available meanings and therefore has the properties of an ambiguous word (e.g., Bouton & Bolles, 1985): The performance

it evokes depends on which meaning is retrieved by the context.

My students and I have recently emphasized an interesting finding that has almost always emerged in our research on contextual control of conditioned responding. Conditioning performance is much less sensitive to the context than is extinction performance. Thus, as I suggested above, in any of several conditioning preparations, a context switch after conditioning appears to have little effect on the conditioned response (Bouton & King, 1983; Bouton & Peck, 1989; Brooks & Bouton, in press; Grahame et al., 1990; Hall & Honey, 1989; Kaye, Preston, Szabo, Druiff, & Mackintosh, 1987). Conditioned responding thus transfers well across contexts. However, a switch out of the extinction context almost always causes a loss of extinction performance (e.g., Bouton & Bolles, 1979a; Bouton & King, 1983; Bouton & Peck, 1989; Bouton & Ricker, in press; Bouton & Swartzentruber, 1989). From a retrieval perspective, memories of extinction appear to be especially dependent on the context for retrieval. In effect, conditioning performance can recover after extinction because the animal forgets about extinction when it is removed from the extinction context.

This emphasis has led to a relatively novel perspective on spontaneous recovery. The passage of time may create changes in both internal and external background cues (e.g., Bouton, 1993; Spear, 1978). As a result, manipulation of a retention interval introduces a change of temporal context. If retention intervals work as context switches, then performance that seems to be lost with a change of physical context should likewise be lost over time; similarly, performance that is not lost with a context switch likewise should not be lost over time. So far, the parallel has held (Bouton, 1993). Conditioned excitation is often unaffected by long retention intervals (e.g., Hendersen, 1985), just as it is relatively unaffected by a physical context switch. In contrast, extinction, which is lost so readily with a physical context switch, is also readily lost over time. The idea that the passage of time provides a change of context suggests that spontaneous recovery is just another renewal effect that happens when the subject is moved out of the temporal extinction context.

The hypothesis, then, is that renewal and spontaneous recovery both are due to a failure to retrieve the memory of extinction. In either case, the subject forgets extinction outside the extinction context. If this were true, then presenting a cue that retrieves extinction at the time of the test should reduce the recovery of responding. Brooks and Bouton (1993) showed that spontaneous recovery can indeed be attenuated by presentation of an extinction cue that had preceded some of the extinction trials. The renewal effect caused by returning the rat to the conditioning context after extinction in another context is likewise attenuated by an extinction cue (Brooks & Bouton, in press). In both sets of experiments we found no evidence that the extinction cue attenuated test performance because it unconditionally disrupted performance to the CS or because it had acquired direct excitatory or inhibitory associations with the US. Instead, the cue appeared to remind the animal of extinction. Spontaneous recovery and renewal may result from a com-

mon mechanism—a failure to retrieve what was learned during extinction.

An emphasis on memory retrieval thus integrates the effects of context and retention interval in extinction. It also promises to integrate extinction with a number of other interference paradigms in which the CS is associated with different outcomes in different phases of the experiment (Bouton, 1993). In all such paradigms, contextual cues and the passage of time play important roles in determining performance. For example, consider counterconditioning, the paradigm in which a CS first is paired with one US (e.g., shock or food) and then with a qualitatively different US (e.g., food or shock) in a second phase. Phase 2 performance, of course, eventually replaces Phase 1 performance. However, that performance may be as sensitive to context and time as is extinction performance. My students and I have recently observed spontaneous recovery (Bouton & Peck, 1992), the renewal effect (Peck & Bouton, 1990), and reinstatement (Brooks, Hale, Nelson, & Bouton, 1994) after counterconditioning. That is, the Phase 1 response recovers (and the Phase 2 response becomes suppressed) with manipulations of retention interval or physical context. Once again, information from both phases appears to be retained, and performance is influenced by time and context. Time and context are similarly important in other paradigms, such as discrimination reversal learning (Bouton & Brooks, 1993; Spear et al., 1980; Thomas, McKelvie, Ranney, & Moye, 1981), latent inhibition (e.g., Channell & Hall, 1983; Hall & Channell, 1985; Kraemer & Roberts, 1984), and, of course, human verbal interference (e.g., Greenspoon & Ranyard, 1957; Underwood, 1948). Conditioning theory has traditionally used different mechanisms to explain the different paradigms, although it has almost always emphasized learning or storage deficits (see Bouton, 1993). Acknowledging the general role of memory retrieval ties the paradigms together in a single, unified framework (Bouton, 1993; see also Spear, 1981) and helps explain the context, time, and recovery effects that appear to occur in all of them.

Remembering and forgetting are important in Pavlovian conditioning. In extinction and a number of other interference paradigms, memories of both phases are retained, and performance depends on which is remembered or retrieved. This approach to conditioning highlights two venerable sources of forgetting: Retrieval failure caused by the absence of contextual retrieval cues (e.g., Tulving, 1974) and interference caused by the learning of conflicting information (e.g., McGeoch, 1932). The understanding of extinction and other paradigms is enriched by a memory perspective. Conversely, an understanding of memory may be enriched by putting it in a conditioning perspective. In the next section, I will consider this possibility further.

How Are Retrieval Cues Learned?

Associations in Behavior and Memory

Memory has actually been a part of conditioning theory since at least the mid-1970s. At about that time, the field of

learning began to accept the idea that a CS elicits conditioned responding because it retrieves a memory representation of the US. Rozeboom (1958) argued that stimulus-stimulus theories of learning suggest a novel empirical fact about conditioning: If the response to the US were changed after conditioning, then any corresponding change in the response evoked by the CS would indicate that the CS controls behavior through its association with the US.

The crucial data came about 15 years later. Rescorla (1973) conditioned fear of a light CS in rats by pairing it with a loud noise US. After conditioning occurred, he repeatedly presented the noise alone in order to habituate the rats' fear of it. When the light CS was then tested on its own, Rescorla found that the noise habituation treatment had reduced fear of the light that signaled it. He suggested that habituation had modified the rats' memory representation of the noise. When the light was now presented again, it retrieved a memory of a less frightening event. Complementary experiments showed that exposure to stronger USs after conditioning analogously increased the rats' fear of the CS (Rescorla, 1974b). In the mid-1970s, then, Rescorla wrote of the CS activating a US representation, or "image," as a result of conditioning (Rescorla, 1974a). US revaluation procedures have produced similar results in a wide variety of conditioning preparations (Holland & Rescorla, 1975; Holland & Straub, 1979; Holloway & Domjan, 1993; Rashotte, Griffin, & Sisk, 1977). They work in operant conditioning as well (e.g., Adams & Dickinson, 1981; Colwill & Rescorla, 1986). Pavlovian CSs and instrumental actions are now almost routinely viewed as events that are associated with memory representations of USs or reinforcers (for a somewhat different perspective on operant learning, see Dickinson, 1989; Dickinson & Balleine, 1994).

The idea that the CS activates or retrieves a US representation was subsequently adopted by nearly every theory of conditioning. Wagner's (e.g., 1978) short-term memory model, which described conditioning within an information-processing system familiar to cognitive psychologists (i.e., Atkinson & Shiffrin, 1968), emphasized the CS's role as a cue that retrieves a memory of the US from long-term memory into short-term memory. Although other models differ in important ways, they have usually been viewed within a similar memory framework (e.g., Pearce, 1987; Pearce & Hall, 1980). A more recent extension of Wagner's theory (e.g., Wagner, 1981; Wagner & Brandon, 1989) puts conditioning in a more connectionist memory architecture. The CS and US are represented as nodes; activation transfers between nodes along the associations that are learned between them. As before, however, the CS arouses behavior because it activates a memory node that corresponds to the US.

Conditioning is thus viewed as more than a procedure that merely permits new stimuli to control behavior; it engages a learning process that allows retrieval cues to become associated with to-be-remembered events. CSs are examples of retrieval cues. Classical conditioning research investigates how they acquire their power. Theories of conditioning should be of interest even to researchers whose major interest is memory rather than learning.

Recent research on conditioning has unquestionably produced insights into how organisms learn about retrieval cues. The learning of the CS-US association is now thought to follow an error-correction rule in which the association is changed on each trial so as to reduce the difference between what the cue predicts and what actually occurs on the trial (Bush & Mosteller, 1955; Mackintosh, 1975; Pearce, 1987; Pearce & Hall, 1980; Rescorla & Wagner, 1972). The crucial insight of the Rescorla-Wagner model (Rescorla & Wagner, 1972) was that the error that is corrected is the difference between the trial outcome and that predicted by all of the cues present on the trial combined. As any student of conditioning knows, the principle has striking implications. The pairing of two stimuli is not sufficient to cause learning if another cue present on the trial already predicts the target event (e.g., Kamin, 1969). In fact, depending on the associative strength of other cues present on the trial, the pairing of a CS and US can lead to an increment in their association, no change in the association, a possible decrement in their association, or even an inhibitory association between them (e.g., Kremer, 1978; Mackintosh & Cotton, 1985). Unfortunately, the Rescorla-Wagner error-correction rule also implies unlearning when no outcome occurs on trials when one is predicted. In fact, the Rescorla-Wagner learning rule has a number of well-known problems (see Miller, Barnett, & Grahame, 1993, for one recent review). Nonetheless, the conditioning literature that it helped stimulate in the 1970s and 1980s is essential reading for anyone looking for a fine-grained analysis of how retrieval cues acquire their power.

Several writers have noted that many ideas embedded in conditioning theories are also built into a number of connectionist, or parallel-distributed-processing, models of human memory (e.g., Baker, Mercier, Valle-Tourangeau, Frank, & Pan, 1993; Gluck, 1992; Gluck & Bower, 1988; Kehoe, 1988; Shanks, 1991). Such models often resemble conditioning theory taken to an extreme. The world is full of features, represented by nodes, that become interassociated through experience. A dog, for example, is a collection of many features (e.g., it is black, it has a tongue, it has a tail, it has four legs), and the nodes corresponding to those features are associated when they co-occur (cf. McClelland & Rumelhart, 1985). The connections between nodes strengthen when the nodes are jointly activated and weaken when they are not; they can also take on either positive (excitatory) or negative (inhibitory) values. Activation of one node will excite other nodes according to the strength (or "weight") of the connection. Networks organized along these lines can learn and do impressive things. They can generalize, discriminate, categorize, and classify; presentation of a new dog (a unique new collection of features) can excite enough nodes so that the whole net of connected nodes is activated, and the system can respond appropriately. Performance also falls apart "gracefully" as the input becomes increasingly distorted or if some of the nodes or their connections die (e.g., McClelland, Rumelhart, & Hinton, 1986).

Connectionist models interest conditioning researchers. The rules that govern how the connections change with

experience are by definition learning rules; in an interesting case of convergent evolution, at least some of them use a version of the Rescorla-Wagner model known as the delta rule or the least mean squares rule (Widrow & Hoff, 1960). Connectionist models are also consistent with the spirit of traditional learning theory, because they build complex cognitive processes (e.g., categorization, pattern recognition) up from the bottom without invoking awkward executive decisionmakers or homunculi whose own behavior demands further explanation. Indeed, the approach even appeals to some radical behaviorists (Donahoe, Burgos, & Palmer, 1993; Palmer & Donahoe, 1992).

Surprising things can happen, however, when a system learns large sets of associations in parallel. One almost irresistible aspect of connectionist models is that unexpected things can emerge from the computer simulations. A back-propagation network (e.g., Rumelhart, Hinton, & Williams, 1986) can require a surprisingly large number of trials to simulate a simple conditioning experiment. In many of the models, the network can cope admirably with a set of problems that it is given in one phase, but if new problems are trained in a second phase the outputs associated with Phase 1 are destroyed (e.g., McCloskey & Cohen, 1989). This *catastrophic interference* problem was foreshadowed and addressed by available research on extinction and other interference paradigms in conditioning. Although some models may be less affected by it (e.g., Kruschke, 1992; Sloman & Rumelhart, 1992), there may be a general lesson here. If the goal is a network that will learn and remember in a realistic way, then one must start with a clear vision of what it is that is realistic. The conditioning literature can play a role here. Research on conditioning provides an extensive and systematic database that constrains the kinds of things that networks should and should not do.

The idea of parallel and distributed associations can reciprocally enrich the understanding of what is learned in simple conditioning experiments. For example, the CS, US, and context may be rich, multifaceted stimuli that require numerous interconnections among their many features to be represented in memory; animals may need to learn about complex stimuli, and that learning may be inherently associative (e.g., Hall, 1991; McLaren, Kaye, & Mackintosh, 1989). The context may be especially complex; interestingly, rats appear to need some time in a conditioning box before they can associate it with shock (e.g., Fanselow, 1986), perhaps because they need some time to form an adequate representation of it. In a standard conditioning experiment in which a CS and US are paired within a context, multiple parallel associations are further possible between each of the nominal stimuli. In addition to the focal CS-US association, animals learn context-CS associations (Rescorla, 1984), CS-context associations (Marlin, 1982), CS-CS associations (i.e., within-compound associations, e.g., Rescorla & Durlach, 1981), and context-US associations (e.g., Baker, Mercier, Gabel, & Baker, 1981; Bouton & King, 1983). The evidence from operant conditioning suggests an equally rich set of associations there (e.g., Colwill & Rescorla, 1986; Rescorla, 1991). Even a deliberately simplified conditioning experiment thus provides an

opportunity to store a complex memory composed of multiple features and associations. The conclusion seems to justify the multiple-attribute approach to animal memory long advocated by Spear (e.g., 1973, 1978; see also Bower, 1967, and Underwood, 1969). Simple conditioning may result in a complex set of connections between the stimuli that are part of the overall experience.

Is there any justification for this emphasis on associations? Although other views of what is learned in conditioning experiments are available (e.g., Dickinson, 1989; Gallistel, 1990), the association-based approach has provided a relatively parsimonious explanation of an astonishing amount of data. Furthermore, recent research in the neurobiology of learning and memory has provided support for the idea that learning facilitates synaptic transmission—the connection, or association, between two neurons (e.g., Fanselow, 1993). In long-term potentiation, for example, one neuron's ability to activate a second neuron increases when activation of the first is paired with activation of the second (e.g., Brown, Chapman, Kairiss, & Keenan, 1988). Interference with the biochemical processes known to mediate this phenomenon can also interfere with classical conditioning (e.g., Kim, DeCola, Landeira-Fernandez, & Fanselow, 1991). In invertebrates, associative learning procedures also enhance transmission across synapses through mechanisms that are becoming increasingly understood (see, e.g., Byrne & Crow, 1991; Hawkins, Abrams, Carew, & Kandel, 1983). There is an exciting convergence of ideas about conditioning, connectionism, and the brain in some physiological approaches to the learning and memory processes that are represented in conditioning (e.g., Gluck & Myers, 1993; Schmajuk & DiCarlo, 1992).

Hierarchical Structures in Behavior and Memory

Recent research in conditioning suggests, however, that there may be other products of learning besides the simple binary association. As I discussed earlier, contexts often do not appear to control performance to CSs through their direct association with the US (see also, e.g., Bonardi, Honey, & Hall, 1990; Hall & Honey, 1989, 1990; Honey, Willis, & Hall, 1990). Earlier I suggested that the context might retrieve the CS's meaning, or current association with the US. Such a view implies a hierarchical structure to behavior and memory (see also Holland, 1992; Rescorla, 1991).

Hierarchical structure has also been invoked to explain the behavioral phenomenon known as *occasion setting* (see Holland, 1992, for a review). Occasion setting can emerge in experiments in which animals are required to learn simple discriminations. For example, in feature-positive discriminations, a "feature" CS marks trials on which another CS (the "target" CS) is reinforced (as opposed to not reinforced). In feature-negative discriminations, the feature conversely marks trials on which the target is not reinforced. In either case, the animals learn to respond on reinforced trials and to withhold responding on nonreinforced trials. Either type of discrimination is easy to explain with simple binary

associations (e.g., Rescorla & Wagner, 1972); the positive feature should acquire an excitatory association with the US, and the negative feature should acquire an inhibitory association. However, in an extensive series of experiments on these and other discrimination problems, Holland has shown that if the feature precedes the target CS in time during training, or if it is simultaneous with the target but is much less salient, then it may control performance to the target by signaling the target-US (or target-no US) association. The evidence generally reduces to the fact that the feature's ability to control responding is independent of its binary association with the US. For example, a negative occasion setter (a feature that marks nonreinforced trials) does not have the properties expected of cues with simple inhibitory associations. It is readily converted into a conditioned excitor (i.e., it shows no retardation in the acquisition of an excitatory CS-US association; cf. Rescorla, 1969), and it does not lose its ability to turn off responding to the target after it has been converted into one (Holland, 1984). Also, unlike a conditioned inhibitor, it does not inhibit performance to a separately trained target except when the new target has itself been in a similar occasion-setting discrimination. A positive occasion setter, analogously, may turn on responding to the target CS even after extensive extinction exposure to it alone (Holland, 1989a; Rescorla, 1986). Contexts may have similar properties (see Bouton & Swartzentruber, 1986; Swartzentruber, 1991; Swartzentruber & Bouton, 1988). There is no complete agreement on how occasion setters operate, but they seem to activate a whole CS-US association rather than a CS node or a US node alone.

Although hierarchical organization has long been emphasized in human memory (e.g., Estes, 1973, 1976), I am not aware of a connectionist architecture that can handle the basic facts of occasion setting. It is easy to simulate a simple discrimination; the main constraint is that the occasion setter must be shown to operate independently of behavioral effects that would suggest control by its direct association with the US or CS. An occasion setter may be put into an associationistic framework by assuming that it becomes associated with the unit comprised of the CS-US (or CS-no US) association. A connection between the occasion setter and the CS-US association could be strengthened according to the Rescorla-Wagner delta rule; in this case, the outcome is the CS-US unit instead of the US on its own (Rescorla, 1990; see also Bonardi, 1991). Although this proposal has appeal, putting the occasion setter in this sort of structure does not solve one of the major puzzles in the occasion setting literature: Under some circumstances, an occasion setter may *transfer* and set the occasion for a completely different CS-US unit (e.g., Holland, 1989a, 1989b; Holland & Lamarre, 1984; Lamarre & Holland, 1987). The problem is perhaps waiting for a new collaboration between conditioning and memory modelers. Whatever the occasion setting structure looks like, and however one supposes it is learned, it has implications for mechanisms of memory retrieval. Discrimination problems, as well as interference problems such as extinction and counterconditioning, may boil down to hierarchical memory problems.

An alternative to a hierarchical structure might be to assume control by a configural cue. That is, any combination of cues could produce its own unique stimulus, and the combination could be associated directly with the US (e.g., Pearce, 1987). The view runs into trouble accommodating some of the facts surrounding occasion setting and discrimination learning (see Holland, 1991, and Bouton & Nelson, 1994, for more detailed discussions). Furthermore, occasion setting and contextual control effects are not merely the result of unique cues arising from *perceptual* interactions among the stimuli (e.g., Bouton, 1993; Holland, 1992). Some human memory models (e.g., Humphreys, Bain, & Pike, 1989) avoid this difficulty by positioning the configural cue within the hypothetical memory system rather than at stimulus input (the perceptual level). A general challenge for such theories, however, will be to explain why cues and contexts can transfer and affect retrieval of memories that have not been explicitly encoded with them (e.g., Holland, 1989a, 1989b, 1991; Holland & Lamarre, 1984; Lamarre & Holland, 1987; Swartzentruber & Bouton, 1988).

To summarize, it is apparent that conditioning and memory theorists are working on overlapping, if not identical, problems. Classical conditioning can be viewed as a method for studying the fundamental mechanisms involved in memory retrieval. There is an opportunity for synergy here; discoveries in one field can help the other discover more. However, even this expanded perspective on the mechanisms of conditioning and remembering is too narrow. For example, it does not address one of the most interesting puzzles uniquely highlighted by the retrieval analysis of extinction. Not all memories acquired in conditioning experiments are remembered equally well; extinction seems more readily forgotten than conditioning. To address this issue, it is necessary to consider the functional context in which conditioning and remembering occur. This exercise will also further illustrate how learning and memory are inseparable aspects of behavior.

The Functions of Forgetting

There has been a great deal of recent interest in the functional aspects of memory. For example, comparative studies suggest that birds that cache their food can have impressive memory capacities adapted to the task (e.g., Kamil & Balda, 1990; Sherry, 1988; Shettleworth, 1990). These adaptations are sometimes correlated with morphological adaptations that have also evolved to deal with food storing. Research on this problem reinforces the importance of learning and memory to animals in their natural environments. However, the emphasis here tends to be on the benefits of remembering. To understand why extinction, as opposed to conditioning, seems so readily lost over context and time, it is useful to recognize that there are also benefits to forgetting.

Anderson and Schooler (1991) provided an insightful start on this problem. They demonstrated that the shape of the familiar forgetting curve, in which information is forgotten in a negatively accelerated fashion over time (e.g.,

Ebbinghaus, 1885/1964; Slamecka & McElree, 1983; Wixted & Ebbesen, 1991), reflects the temporal structure of the environment in which remembering organisms actually operate. They analyzed how items that humans have to remember in day-to-day life are actually distributed over time. For example, words encountered in newspaper headlines (e.g., "Reagan" or "Qaddafi") require the reader to retrieve related information; Anderson and Schooler therefore analyzed 2 years' worth of headlines in the *New York Times*. The probability that a word appeared on a particular day was a lawful function of the number of days since its last occurrence; it was most probable if the word had been presented the day before, and its probability declined as a function of days since the last occurrence in a fashion that perfectly approximated the typical forgetting curve. Similar results were obtained when Anderson and Schooler analyzed words spoken to children and the sources of E-mail messages to Anderson. The forgetting function thus matches the probable distribution of items in time; when items are forgotten, they are lost precisely as their probable usefulness declines. Anderson and Schooler (1991) suggested that "human memory mirrors, with a remarkable degree of fidelity, the structure that exists in the environment" (p. 404).

The forgetting of information as its usefulness declines presumably is beneficial because it reduces information overload; the organism can cope only with a finite number of items at any one time. Miller, Kasrow, and Schachtman (1986) suggested that forgetting older information may be useful in protecting more recent information from interference. They also suggested that retrieval failure is a particularly functional mechanism of forgetting because it keeps the information available (if not currently accessible) for later use should the need arise.

It seems likely, however, that different items, or types of information, can vary widely in how they are actually distributed over time. For example, whereas my memory of what Qaddafi did on January 1, 1986, may not be of much use to me 100 days later, my recollection of who Qaddafi is, where he lives, and what his political beliefs are may be; thus, one may expect that Qaddafi's more stable attributes will be relatively stable in memory over time. If forgetting is adaptively organized, the forgetting function for such attributes should be correspondingly flat. In fact, Anderson and Schooler (1991) also found that the probability that an item would appear again declined at a slower rate if the two prior appearances had been separated in time rather than close together in time—there should be a flatter retention function for items encountered with long rather than short study lags (the trial spacing effect; e.g., Glenberg, 1976). The main difference between items that are remembered over the short as opposed to long term may be their inherent distribution and thus their likely usefulness over different scales of time. There may be no qualitative difference between short-term and long-term memory (see also Stadon, 1988).

It is interesting to note how pervasively important the stability-variability dimension is to the ecology of learning and memory. Learning (and memory) are useful primarily

when an organism needs to adapt to a variable but predictable environment. If the environment did not vary, it would be equally effective, and less costly in terms of expensive brain tissue and energy, for evolution to prepare the animal to respond innately to environmental input (e.g., Alcock, 1993). Learning and remembering thus pay when the environment is variable. Learning plus forgetting pays when environmental variability becomes more extreme; if the environment is truly variable, what is learned today will not be useful for long. Forgetting is functional, especially when the variability of the environment is great. Innate behavior, learned behavior, and learned-then-forgotten behavior occupy ordered positions in evolution's response to increasing environmental variability.

Arnold (1978) analyzed the theoretical benefits of forgetting in a mimicry system in which a hypothetical predator must learn to discriminate between edible Batesian mimics and the nasty, distasteful models they resemble. The classic example of such a system is the tasty viceroy butterfly, which has evolved to look like the nasty-tasting monarch. The evolution of mimicry is a problem in which the interests of learning theorists and behavioral ecologists overlap; as far as I know, the overlap has not been exploited sufficiently. The predator's encounter with a nasty-tasting model is a conditioning trial in which the model's features are associated with sickness or a bitter taste. As a result of learning, the predator avoids the model, and the avoidance generalizes to mimics according to the similarity between the two (e.g., Morrell & Turner, 1970). The predator can benefit from learning about and remembering the model, because it will then not waste resources tracking down and handling another one. However, it can also benefit from forgetting about the model if forgetting increases the likelihood that it will accept and eat (and thus benefit from) a palatable mimic.

With just a few assumptions, Arnold (1978) showed that a predator's memory performance should evolve in a manner that reflects the dispersion of nasty-tasting models and palatable mimics in the environment. Memory should evolve so that the predator remembers the model long enough to skip over clumps of models but forgets soon enough so that it will sample (and thus benefit from) the palatable mimics that come between. In the end, an optimal predator will forget more slowly as the model's expected clump size or run length increases. It will also forget more slowly as the mimic's clump size decreases; as mimics become more rare, the predator encounters longer runs of models separated by just a few mimics. A perfect memory (in which the predator avoids all similar items in the future after learning about the nasty-tasting model) is rarely optimal; it pays only in environments in which the distribution of models and mimics is extremely "coarse-grained" such that the prey type is constant within a predator's lifetime but is relatively unpredictable from generation to generation. (If it were predictable between generations, learning would not be necessary, and model avoidance could become innate.) Arnold's analysis also revealed a payoff for multiple-trial, rather than one-trial, learning. When models are distributed such that there are both large clumps and strays, multiple-

trial learning allows the predator to distinguish between the clumps and the strays.

Arnold's (1978) theoretical work is a nice illustration of how learning and forgetting are intimately related to one another and to similar ecological factors. It is interesting to note that this sort of ecological approach lends support to the idea that physical context and time have similar effects (Bouton, 1993). Nasty-tasting models and palatable mimics can be dispersed in two ways—in space or over time. Forgetting over either dimension presumably leads to the same cost-benefit analysis.

Although Arnold (1978) did not consider it, his analysis of a predator's memory of the nasty-tasting model can be readily extended to its memory of the palatable mimic. This extension is interesting, because when a predator captures and eats a mimic after conditioning, one has the equivalent of an extinction (or perhaps a counterconditioning) trial. (Recall that counterconditioning is similar to extinction in that the CS's second meaning also appears to be forgotten rapidly over space and time; e.g., Bouton & Peck, 1992; Peck & Bouton, 1990). Forgetting of the mimic would have the same general payoffs as forgetting of the model. As before, the predator should forget extinction slowly if the run length or clump size of mimics is large. It should also forget extinction slowly if the model's expected clump size is small (this implies that nasty-tasting models are rare).

Taken together with some additional ideas, this analysis might suggest a payoff for the predator that forgets extinction more readily than conditioning. The actual dispersal of models and mimics in the environment is not random; it is controlled by its own selection pressures. Mimicry evolves because it allows the mimic to avoid being eaten. If predators learn, the mimic benefits only when there are enough nasty-tasting models in the environment to provide the requisite conditioning trials. For this reason, mimics should not be too numerous relative to models (e.g., Brower, 1960; Edmunds, 1974; Turner, 1977). Thus, mimics might be distributed in smaller clumps than are models. The models themselves might benefit from the same distribution, because there are fewer extinction opportunities that would jeopardize their own defense. If mimics are distributed in smaller clumps, they should be forgotten relatively quickly. A predator foraging in such an environment would behave adaptively if it forgot extinction more readily than conditioning. In this arrangement, extinction should be forgotten more readily than conditioning over space and time.¹

Is this payoff for faster forgetting of extinction a peculiar characteristic of predators foraging on models and mimics? The argument would hold whenever extinction trials are distributed in relatively isolated temporal or spatial pockets.

¹ Predicting the model's and mimic's actual distribution is complicated. There is evidence from at least one laboratory experiment to suggest that mimics may benefit even when they are more numerous than models (Brower, 1960). If this meant that mimics in the field were often distributed in larger clumps than models, it would weaken the argument. However, the result is not convincing. Furthermore, models would suffer from an overabundance of mimics and presumably would evolve so as to prevent this.

Many natural reinforcers may be distributed in such a way. Consider flower nectar. For a bee foraging among flowers, nectar provides a US that allows the bee to identify and track the flowers by their visual and olfactory features (e.g., Bitterman, 1988; J. L. Gould, 1984). Nectar also provides an operant reinforcer for particular behaviors; this presumably allows bees to specialize on morphologically diverse flowers (Heinrich, e.g., 1976). If the bee forages to depletion, it will undergo extinction, but the nectar in a flower will be renewed again after a period of time. For the bee, therefore, forgetting of extinction over time pays. In addition, similar flowers located in other patches may contain nectar even after flowers in the first patch have been emptied. Forgetting of extinction thus will also pay when an animal moves to a new physical context. Correlations between cues and consequences may often renew themselves with a change of context or time. There may often be a payoff for forgetting extinction.

There are still other reasons to forget extinction more readily than conditioning. A conditioning trial provides an opportunity for an animal to sample from, and make inferences about, the rest of the world (e.g., Staddon, 1988). If the world is composed of at least two types of trials (say, models and mimics), the probability of sampling a particular type on any trial will reflect its actual incidence in the population. Learning and memory may have evolved to take advantage of that fact. If one type is truly more abundant than the other, then the odds are good that the animal will encounter it first. An early run of trials of one type (conditioning) would reflect a dominance of that item in the population. Subsequent samples of other types, if not too numerous, would be mere exceptions to the rule. In this sense, treating the second type of trial as less probable than the first presumably pays; it is an adaptation to a statistical feature of sampling items from real populations. Memory may be designed to treat second-learned information as an exception to the rule. Extinction (or any other second association connected with a CS; e.g., Bouton, 1993) should be ignored or forgotten with changes of place or time.

Another factor that could influence forgetting rate is the significance, or value, of the US (Bouton, 1993). It seems plausible that especially important things may be less susceptible to forgetting. For instance, it could be extremely costly for an animal to forget a traumatic learning experience with a temporal or physical context switch—one mistake and nature's experiment is over. It is worth noting that much of what is known about remembering and forgetting, particularly in humans, is based on experiments conducted on memory for extremely trivial information (Henderson, 1985). Words, sentences, and *New York Times* headlines are not as significant as many of the events that are experienced outside of the lab. There is reason to believe that particularly aversive or nasty experiences should be remembered well. Arnold's (1978) theoretical modeling of the mimicry system suggested a payoff for forgetting nasty items more slowly than less nasty items. Perhaps this contributes to why excitation, the memory for a CS's connection with a significant US, is remembered better than extinction, the memory for the US's absence.

In fear conditioning, the relative stability of excitation and instability of inhibition suggests that memory in the fear system may be organized to make the animal "adaptively conservative" over time (Henderson, 1985; Mineka, 1992). If becoming wary and conservative pays, perhaps this is why fear inhibition is forgotten more readily than fear excitation. Attractive as this idea is, it does not stand up to existing data. Henderson (1985) rejected it because of his own evidence suggesting that old fears are more malleable than newer fears. Fear was conditioned in rats with a medium-intensity shock and then tested 1 or 60 days later. If the rat received two weaker shocks just before the test, it decreased its fear of the CS (cf. Rescorla, 1973) at a 60-day retention interval, but not at the 1-day interval. The rats with the older fear were readier to downgrade it—the opposite of being more warily conservative. (They also were readier to increase their fear if exposed to two stronger shocks.) In my own laboratory, my colleagues and I have consistently found parallels between fear conditioning and appetitive (food) conditioning that also question the idea that rats become more wary over time. As in fear conditioning, appetitive excitation is stable over physical context and time and inhibition is less so (Bouton & Nelson, 1994; Nelson, 1994). Renewal effects after extinction, and, of course, spontaneous recovery, are also readily observed in appetitive conditioning (e.g., Bouton & Peck, 1989; Brooks & Bouton, 1993, in press; Pavlov, 1927). Do rats in food conditioning experiments become more wildly optimistic, instead of warily conservative, over space and time? There is no evidence that fear is qualitatively different from appetitive conditioning in how it is remembered and forgotten.

To summarize, it is useful to recognize that learning, remembering, and forgetting all occur in the same biological context. A functional analysis reveals that forgetting confers a benefit, just as learning and remembering do, and it yields hypotheses about why extinction may be forgotten so readily over space and time. Knowledge of the functions of learning and remembering enriches one's appreciation of them. Similarly, knowledge of how animals learn and forget can enrich one's appreciation of adaptation, as is apparent in the evolution of mimicry systems (e.g., Speed, 1993). Conditioning, remembering, and forgetting are interconnected aspects of how animals adapt to their environments.

Extinction, Forgetting, and Relapse

The retrieval analysis of extinction, embellished now with input from the fields of memory and ecology, also has implications for a more applied behavioral science. Extinction (or counterconditioning, or both) is embedded in most cognitive-behavioral therapies that are designed to eliminate unwanted emotions, behaviors, or cognitions. Behavior therapies such as systematic desensitization and implosive therapy were specifically modeled on extinction and counterconditioning, but nearly any kind of therapy can be viewed as a retroactive interference procedure in which new learning is designed to replace old.

The main implication of the basic research on extinction and counterconditioning is that their effects are not permanent. Each leaves the original learning at least partly intact and ready to recover with manipulations of place or time. Each of the laboratory phenomena that support this conclusion—for example, reinstatement, renewal, and spontaneous recovery—are potential mechanisms of relapse (e.g., Bouton & Swartzentruber, 1991). Thus, a switch out of the temporal or physical context or a return to the original conditioning context might cause a recovery of a performance that was eliminated by therapy. The same switches and returns could also undo the effects of natural extinction that might follow traumatic learning and could thus contribute to the seemingly spontaneous emergence of fear or anxiety. Performance after extinction and counterconditioning can return through any of a large number of context manipulations that have been studied in animal laboratories.

The recent description of recovery effects as failures to retrieve memories of extinction has other implications. I have already discussed the data suggesting that spontaneous recovery and renewal are reduced when a subject is given cues that remind it of extinction (Brooks & Bouton, 1993, in press); the implication is that appropriate reminders may help prevent relapse. Therapy reminders may be useful in clinical settings (e.g., E. B. Foa, personal communication, August 20, 1993; George, 1989; Marlatt, 1990). Research on context effects in human memory may complement the work on animal conditioning. For example, Smith (1979) showed that human memory for a list of words, impaired by a change of context, can be restored if the subject is instructed to merely remember the original context. Analogous instructions could be built into therapy. Context specificity of the memory for words can also be reduced if the words are studied in multiple contexts to begin with (Smith, 1982); the implication is that therapy conducted in several contexts may generalize more effectively. Such shifts in context may retard the rate of acquisition, but the learning that is achieved may be more permanent. Indeed, Schmidt and Bjork (1992) reviewed evidence suggesting that variability in practice can lead to better retention even when it harms acquisition performance. It is not yet clear that these principles apply to the kinds of emotional memories that are investigated in animals and are likely to be a problem clinically. However, a framework is now in place that helps encourage an integration of the separate literatures.

There are other implications of the emphasis on memory retrieval. If a therapist found that a client's fear had recovered in a certain context, one approach to the problem might be to expose the client to the context so as to extinguish it. If the context worked through its direct association with the US (as it does in the reinstatement paradigm), the treatment would weaken the context's ability to excite fear of the CS. However, if the context operates as a hierarchical memory cue (e.g., as an occasion setter), simple exposure would not be sufficient to reduce its influence (e.g., Bouton & Peck, 1989; Bouton & Swartzentruber, 1986; Holland, 1989a; Rescorla, 1985, 1986). Instead, the client would need to be exposed to a CS without the US in that context in order to reduce the context's power (Rescorla, 1986). More infor-

mation is needed on how to reduce the effects of occasion setters and hierarchical retrieval cues.

It is also interesting to consider the previous analysis of the functions of forgetting. The fact that it may be functional to forget extinction more readily than conditioning puts the clinical problem in a new light: Relapse may occur because it results from mechanisms that ordinarily are adaptive. Fear, for example, has a well-known biological function; it organizes a constellation of physiological and behavioral responses that help the organism defend itself (e.g., Bolles & Fanselow, 1980; Hollis, 1982). Extinction is also functional, of course, because fear responses are energetically costly, and it would not pay for them to persist in the face of new environmental contingencies. However, the occurrence of relapse is in turn a functional hedge against the possible distribution of events in the world—the extinguished response recovers because the behavior may be needed again. Relapse is to be expected from what is known about both the mechanisms of extinction and the apparent design of the memory system.

When anxiety and fear are concerned, humans may follow the principles of adaptive conservatism, as Mineka (1992) recently argued. The generalization of fear may increase over time (e.g., Riccio, Richardson, & Ebner, 1984); organisms may forget attributes of the US that effectively make them respond with fear to a wider range of situations (Henderson, Patterson, & Jackson, 1980); fear excitation is remembered well over time, whereas fear inhibition is not; fear extinction is inherently context specific. However, such phenomena, as I have already suggested, are not unique to the fear system. Fear and appetitive learning are not necessarily different. Perhaps they both have adapted to the universal constraints of how the world is organized. Alternatively, either one could have merely co-opted learning mechanisms that were adapted to a different situation; this is the evolutionary principle of *exaptation* (S. J. Gould & Vrba, 1982; see also Sherry & Schacter, 1987). One consequence of accepting the more general position, of course, is that it broadens the range of situations over which context and retrieval mechanisms can be expected to contribute. Relapse (renewal, reinstatement, and spontaneous recovery) will happen not only after fear or anxiety extinction but also after the extinction of learning based on positive reinforcers. One example of the latter situation is drug abuse, which is often modeled on principles of positive reinforcement (e.g., Stewart, de Wit, & Eikelboom, 1984; Stoleran, 1992), and where relapse is a well-known problem (e.g., Marlatt & Gordon, 1985).

Conclusion

Extinction, one of the oldest and most fundamental problems in conditioning and behavior theory, is richly connected with issues in human memory, comparative cognition, behavioral ecology, and behavior therapy. My main goal has been to illustrate the usefulness of combining learning and memory concepts. However, exploring the connections between conditioning theory and each of these

other areas can enrich the understanding of conditioning phenomena. Conditioning theory can in turn contribute to these other areas. Behavioral science in general benefits if the bridges among its subdisciplines are kept open and busy.

References

- Adams, C. D., & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Quarterly Journal of Experimental Psychology*, 33B, 109–122.
- Alcock, J. (1993). *Animal behavior: An evolutionary approach* (5th ed.). Sunderland, MA: Sinauer.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.
- Arnold, S. J. (1978). The evolution of a special class of modifiable behaviors in relation to environmental pattern. *The American Naturalist*, 112, 415–427.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation* (Vol. 2, pp. 89–105). San Diego, CA: Academic Press.
- Baker, A. G., Mercier, P., Gabel, J., & Baker, P. A. (1981). Contextual conditioning and the US preexposure effect in conditioned fear. *Journal of Experimental Psychology: Animal Behavior Processes*, 7, 109–128.
- Baker, A. G., Mercier, P., Valle-Tourangeau, F., Frank, R., & Pan, M. (1993). Selective association and causality judgments: Presence of a strong causal factor may reduce judgments of a weaker one. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 414–432.
- Bitterman, M. E. (1988). Vertebrate–invertebrate comparisons. In H. J. Jerison & I. Jerison (Eds.), *Intelligence and evolutionary biology* (pp. 251–276). Berlin: Springer-Verlag.
- Bolles, R. C., & Fanselow, M. S. (1980). A perceptual–defensive–recuperative model of fear and pain. *Behavioral and Brain Sciences*, 3, 291–323.
- Bonardi, C. (1991). Blocking of occasion setting in feature-positive discriminations. *Quarterly Journal of Experimental Psychology*, 43B, 431–448.
- Bonardi, C., Honey, R. C., & Hall, G. (1990). Context specificity of conditioning in flavor-aversion learning: Extinction and blocking tests. *Animal Learning & Behavior*, 18, 229–237.
- Bouton, M. E. (1984). Differential control by context in the inflation and reinstatement paradigms. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 56–74.
- Bouton, M. E. (1991). Context and retrieval in extinction and in other examples of interference in simple associative learning. In L. Dachowski & C. F. Flaherty (Eds.), *Current topics in animal learning: Brain, emotion, and cognition* (pp. 25–53). Hillsdale, NJ: Erlbaum.
- Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, 114, 80–99.
- Bouton, M. E., & Bolles, R. C. (1979a). Contextual control of the extinction of conditioned fear. *Learning and Motivation*, 10, 445–466.
- Bouton, M. E., & Bolles, R. C. (1979b). Role of conditioned contextual stimuli in reinstatement of extinguished fear. *Journal of Experimental Psychology: Animal Behavior Processes*, 5, 368–378.
- Bouton, M. E., & Bolles, R. C. (1985). Contexts, event-memories, and extinction. In P. D. Balsam & A. Tomie (Eds.), *Context and learning* (pp. 133–166). Hillsdale, NJ: Erlbaum.
- Bouton, M. E., & Brooks, D. C. (1993). Time and context effects on performance in a Pavlovian discrimination reversal. *Journal of Experimental Psychology: Animal Behavior Processes*, 19, 165–179.
- Bouton, M. E., & King, D. A. (1983). Contextual control of the extinction of conditioned fear: Tests for the associative value of the context. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 248–265.
- Bouton, M. E., & Nelson, J. B. (1994). Context-specificity of target versus feature inhibition in a feature-negative discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 20, 51–65.
- Bouton, M. E., & Peck, C. A. (1989). Context effects on conditioning, extinction, and reinstatement in an appetitive conditioning preparation. *Animal Learning & Behavior*, 17, 188–198.
- Bouton, M. E., & Peck, C. A. (1992). Spontaneous recovery in cross-motivational transfer (counterconditioning). *Animal Learning & Behavior*, 20, 313–321.
- Bouton, M. E., & Ricker, S. T. (in press). Renewal of extinguished responding in a second context. *Animal Learning & Behavior*.
- Bouton, M. E., Rosengard, C., Achenbach, G., Peck, C. A., & Brooks, D. C. (1993). Effects of contextual conditioning and unconditional stimulus presentation on performance in appetitive conditioning. *Quarterly Journal of Experimental Psychology*, 46B, 63–95.
- Bouton, M. E., & Swartzentruber, D. (1986). Analysis of the associative and occasion-setting properties of contexts participating in a Pavlovian discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 12, 333–350.
- Bouton, M. E., & Swartzentruber, D. (1989). Slow reacquisition following extinction: Context, encoding, and retrieval mechanisms. *Journal of Experimental Psychology: Animal Behavior Processes*, 15, 43–53.
- Bouton, M. E., & Swartzentruber, D. (1991). Sources of relapse after extinction in Pavlovian and instrumental learning. *Clinical Psychology Review*, 11, 123–140.
- Bower, G. H. (1967). A multi-component theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1, pp. 229–325). San Diego, CA: Academic Press.
- Brooks, D. C., & Bouton, M. E. (1993). A retrieval cue for extinction attenuates spontaneous recovery. *Journal of Experimental Psychology: Animal Behavior Processes*, 19, 77–89.
- Brooks, D. C., & Bouton, M. E. (in press). A retrieval cue for extinction attenuates response recovery (renewal) caused by a return to the conditioning context. *Journal of Experimental Psychology: Animal Behavior Processes*.
- Brooks, D. C., Hale, B., Nelson, J. B., & Bouton, M. E. (1994). *Reinstatement after counterconditioning*. Manuscript submitted for publication.
- Brower, J. V. Z. (1960). Experimental studies of mimicry: IV. The reactions of starlings to different proportions of models and mimics. *The American Naturalist*, 94, 271–282.
- Brown, T. H., Chapman, P. F., Kairiss, E. W., & Keenan, C. L. (1988). Long-term synaptic potentiation. *Science*, 242, 724–728.
- Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning*. New York: Wiley.
- Byrne, J. H., & Crow, T. (1991). Examples of mechanistic analyses of learning and memory in invertebrates. In J. L. Martinez, Jr. & R. P. Kesner (Eds.), *Learning and memory: A biological view* (2nd ed., pp. 329–358). San Diego, CA: Academic Press.
- Channell, S., & Hall, G. (1983). Contextual effects in latent inhibition with an appetitive conditioning procedure. *Animal Learning & Behavior*, 11, 67–74.

- Colwill, R. M., & Rescorla, R. A. (1986). Associative structures in instrumental learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 20, pp. 55–104). San Diego, CA: Academic Press.
- Dickinson, A. (1989). Expectancy theory in animal conditioning. In S. B. Klein & R. R. Mowrer (Eds.), *Contemporary learning theories: Pavlovian conditioning and the status of traditional learning theory* (pp. 279–308). Hillsdale, NJ: Erlbaum.
- Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, 22, 1–18.
- Donahoe, J. W., Burgos, J. E., & Palmer, D. C. (1993). A selectionist approach to reinforcement. *Journal of the Experimental Analysis of Behavior*, 60, 17–40.
- Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover. (Original work published 1885)
- Edmunds, M. (1974). *Defence in animals: A survey of anti-predator defences*. Harlow, England: Longman.
- Estes, W. K. (1973). Memory and conditioning. In F. J. McGuigan & D. B. Lumsden (Eds.), *Contemporary approaches to conditioning and learning* (pp. 265–286). New York: Holt, Rinehart, & Winston.
- Estes, W. K. (1976). Structural aspects of associative models for memory. In C. N. Cofer (Ed.), *The structure of human memory* (pp. 31–53). San Francisco: Freeman.
- Fanselow, M. S. (1986). Associative vs. topographical accounts of the immediate shock-freezing deficit in rats: Implications for the response selection rules governing species-specific defensive reactions. *Learning and Motivation*, 17, 16–39.
- Fanselow, M. S. (1993). Associations and memories: The role of NMDA receptors and long-term potentiation. *Current Directions in Psychological Science*, 2, 152–156.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- George, W. H. (1989). Marlatt and Gordon's relapse prevention model: A cognitive-behavioral approach to understanding and preventing relapse. *Journal of Chemical Dependency Treatment*, 2, 125–152.
- Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, 15, 1–16.
- Gluck, M. A. (1992). Stimulus sampling and distributed representations in adaptive network theories of learning. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (Vol. 1, pp. 169–200). Hillsdale, NJ: Erlbaum.
- Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227–247.
- Gluck, M. A., & Myers, C. E. (1993). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus*, 3, 491–516.
- Gould, J. L. (1984). Natural history of honey bee learning. In P. Marler & H. S. Terrace (Eds.), *The biology of learning* (pp. 149–180). Berlin: Springer-Verlag.
- Gould, S. J., & Vrba, E. S. (1982). Exaptation—A missing term in the science of form. *Paleobiology*, 8, 4–15.
- Grahame, N. J., Hallam, S. C., Geier, L., & Miller, R. R. (1990). Context as an occasion setter following either CS acquisition and extinction or CS acquisition alone. *Learning and Motivation*, 21, 237–265.
- Greenspoon, J., & Ranyard, R. (1957). Stimulus conditions and retroactive inhibition. *Journal of Experimental Psychology*, 53, 55–59.
- Hall, G. (1991). *Perceptual and associative learning*. Oxford, England: Clarendon Press.
- Hall, G., & Channell, S. (1985). Differential effects of contextual change on latent inhibition and on the habituation of an orienting response. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 470–481.
- Hall, G., & Honey, R. C. (1989). Contextual effects in conditioning, latent inhibition, and habituation: Associative and retrieval functions of contextual cues. *Journal of Experimental Psychology: Animal Behavior Processes*, 15, 232–241.
- Hall, G., & Honey, R. C. (1990). Context-specific conditioning in the conditioned-emotional-response procedure. *Journal of Experimental Psychology: Animal Behavior Processes*, 16, 271–278.
- Hawkins, R. D., Abrams, T. W., Carew, T. J., & Kandel, E. R. (1983). A cellular mechanism of classical conditioning in *aplysia*: Activity-dependent amplification of presynaptic facilitation. *Science*, 219, 400–405.
- Heinrich, B. (1976). The foraging specializations of individual bumblebees. *Ecological Monographs*, 46, 105–128.
- Henderson, R. W. (1985). Fearful memories: The motivational significance of forgetting. In F. R. Brush & J. B. Overmier (Eds.), *Affect, conditioning, and cognition: Essays on the determinants of behavior* (pp. 43–54). Hillsdale, NJ: Erlbaum.
- Henderson, R. W., Patterson, J., & Jackson, R. L. (1980). Acquisition and retention of control of instrumental behavior by a cue signalling airblast: How specific are the conditioned anticipations? *Learning and Motivation*, 11, 407–426.
- Holland, P. C. (1984). Differential effects of reinforcement of an inhibitory feature after serial and simultaneous feature-negative discrimination training. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 461–475.
- Holland, P. C. (1989a). Feature extinction enhances transfer of occasion setting. *Animal Learning & Behavior*, 17, 269–279.
- Holland, P. C. (1989b). Transfer of negative occasion setting and conditioned inhibition across conditioned and unconditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 15, 183–193.
- Holland, P. C. (1991). Transfer of control in ambiguous discriminations. *Journal of Experimental Psychology: Animal Behavior Processes*, 17, 231–248.
- Holland, P. C. (1992). Occasion setting in Pavlovian conditioning. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 28, pp. 69–125). San Diego, CA: Academic Press.
- Holland, P. C., & Lamarre, J. (1984). Transfer of inhibition after serial and simultaneous feature negative discrimination training. *Learning and Motivation*, 15, 219–243.
- Holland, P. C., & Rescorla, R. A. (1975). The effect of two ways of devaluing the unconditioned stimulus after first- and second-order appetitive conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 1, 355–363.
- Holland, P. C., & Straub, J. J. (1979). Differential effects of two ways of devaluing the unconditioned stimulus after Pavlovian appetitive conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 5, 65–78.
- Hollis, K. L. (1982). Pavlovian conditioning of signal-centered action patterns and autonomic behavior: A biological analysis of function. *Advances in the Study of Behavior*, 12, 1–64.
- Holloway, K. S., & Domjan, M. (1993). Sexual approach conditioning: Tests of unconditioned stimulus devaluation using hormone manipulations. *Journal of Experimental Psychology: Animal Behavior Processes*, 19, 47–55.
- Honey, R. C., Willis, A., & Hall, G. (1990). Context specificity in pigeon autoshaping. *Learning and Motivation*, 21, 125–136.

- Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, 96, 208–233.
- Kamil, A. C., & Balda, R. P. (1990). Spatial memory in seed-caching corvids. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 1–25). San Diego, CA: Academic Press.
- Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. B. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.
- Kaye, H., Preston, G., Szabo, L., Druiff, H., & Mackintosh, N. J. (1987). Context specificity of conditioning and latent inhibition: Evidence for a dissociation of latent inhibition and associative interference. *Quarterly Journal of Experimental Psychology*, 39B, 127–145.
- Kehoe, E. J. (1988). A layered network model of associative learning: Learning to learn and configuration. *Psychological Review*, 95, 411–433.
- Kim, J. J., DeCola, J. P., Landeira-Fernandez, J., & Fanselow, M. S. (1991). N-methyl-D-aspartate receptor antagonist APV blocks acquisition but not expression of fear conditioning. *Behavioral Neuroscience*, 105, 126–133.
- Kraemer, P. J., & Roberts, W. A. (1984). The influence of flavor preexposure and test interval on conditioned taste aversions in the rat. *Learning and Motivation*, 15, 259–278.
- Kremer, E. F. (1978). The Rescorla–Wagner model: Losses in associative strength in compound conditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 4, 22–36.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
- Lamarre, J., & Holland, P. C. (1987). Transfer of inhibition after serial feature negative discrimination training. *Learning and Motivation*, 18, 319–342.
- Lovibond, P. F., Preston, G. C., & Mackintosh, N. J. (1984). Context specificity of conditioning, extinction, and latent inhibition. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 360–375.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276–298.
- Mackintosh, N. J., & Cotton, M. M. (1985). Conditioned inhibition from reinforcement reduction. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 89–112). Hillsdale, NJ: Erlbaum.
- Marlatt, G. A. (1990). Cue exposure and relapse prevention in the treatment of addictive behaviors. *Addictive Behaviors*, 15, 395–399.
- Marlatt, G. A., & Gordon, J. R. (1985). *Relapse prevention: Maintenance strategies in the treatment of addictive behaviors*. New York: Guilford Press.
- Marlin, N. A. (1982). Within-compound associations between the context and the conditioned stimulus. *Learning and Motivation*, 13, 526–541.
- McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114, 159–188.
- McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 3–44). Cambridge, MA: Bradford.
- McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 24, pp. 109–165). San Diego, CA: Academic Press.
- McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, 39, 352–370.
- McLaren, I. P. L., Kaye, H., & Mackintosh, N. J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R. G. M. Morris (Ed.), *Parallel distributed processing* (pp. 102–130). London: Oxford University Press.
- Miller, R. R., Barnet, R. C., & Grahame, N. J. (1993). *Assessment of the Rescorla–Wagner model*. Manuscript submitted for publication.
- Miller, R. R., Kasprow, W. J., & Schachtman, T. R. (1986). Retrieval variability: Sources and consequences. *American Journal of Psychology*, 99, 145–218.
- Mineka, S. (1992). Evolutionary memories, emotional processing, and the emotional disorders. In D. Medin (Ed.), *The psychology of learning and motivation* (Vol. 28, pp. 161–206). San Diego, CA: Academic Press.
- Morrell, G. M., & Turner, J. R. G. (1970). Experiments on mimicry: I. The response of wild birds to artificial prey. *Behaviour*, 36, 116–130.
- Nelson, J. B. (1994). *The effects of a retention interval on feature and target inhibition in a feature negative discrimination*. Unpublished master's thesis, University of Vermont, Burlington, VT.
- Palmer, D. C., & Donahoe, J. W. (1992). Essentialism and selectionism in cognitive science and behavior analysis. *American Psychologist*, 47, 1344–1358.
- Pavlov, I. P. (1927). *Conditioned reflexes*. London: Oxford University Press.
- Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, 94, 61–73.
- Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532–552.
- Peck, C. A., & Bouton, M. E. (1990). Context and performance in aversive-to-appetitive and appetitive-to-aversive transfer. *Learning and Motivation*, 21, 1–31.
- Rashotte, M. E., Griffin, R. W., & Sisk, C. L. (1977). Second-order conditioning of the pigeon's keypeck. *Animal Learning & Behavior*, 5, 25–38.
- Rescorla, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, 72, 77–94.
- Rescorla, R. A. (1973). Effect of US habituation following conditioning. *Journal of Comparative and Physiological Psychology*, 82, 101–106.
- Rescorla, R. A. (1974a). A model of conditioning. In V. S. Rusinov (Ed.), *Mechanism of formation and inhibition of conditional reflex* (pp. 25–39). Moscow: Academy of Sciences of the USSR.
- Rescorla, R. A. (1974b). Effect of inflation of the unconditioned stimulus value following conditioning. *Journal of Comparative and Physiological Psychology*, 86, 101–106.
- Rescorla, R. A. (1984). Associations between Pavlovian CSs and context. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 195–204.
- Rescorla, R. A. (1985). Conditioned inhibition and facilitation. In R. R. Miller & N. E. Spear (Eds.), *Information processing in*

- animals: *Conditioned inhibition* (pp. 299–326). Hillsdale, NJ: Erlbaum.
- Rescorla, R. A. (1986). Extinction of facilitation. *Journal of Experimental Psychology: Animal Behavior Processes*, 12, 16–24.
- Rescorla, R. A. (1990). The role of information about the response–outcome relation in instrumental discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 16, 262–270.
- Rescorla, R. A. (1991). Associative relations in instrumental learning: The Eighteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, 43B, 1–23.
- Rescorla, R. A., & Durlach, P. J. (1981). Within-event learning in Pavlovian conditioning. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 81–112). Hillsdale, NJ: Erlbaum.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning: II. Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
- Riccio, D. C., Richardson, R., & Ebner, D. L. (1984). Memory retrieval deficits based upon altered contextual cues: A paradox. *Psychological Bulletin*, 96, 152–165.
- Rozeboom, W. W. (1958). "What is learned?"—An empirical enigma. *Psychological Review*, 65, 22–33.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 318–362). Cambridge, MA: Bradford.
- Schmajuk, N. A., & DiCarlo, J. J. (1992). Stimulus configuration, classical conditioning, and hippocampal function. *Psychological Review*, 99, 268–305.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3, 207–217.
- Shanks, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 433–443.
- Sherry, D. F. (1988). Learning and adaptation in food-storing birds. In R. C. Bolles & M. D. Beecher (Eds.), *Evolution and learning* (pp. 79–96). Hillsdale, NJ: Erlbaum.
- Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review*, 94, 439–454.
- Shettleworth, S. J. (1990). Spatial memory in food-storing birds. *Philosophical Transactions of the Royal Society, B*, 329, 143–151.
- Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 384–397.
- Sloman, S. A., & Rumelhart, D. E. (1992). Reducing interference in distributed memories through episodic gating. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (Vol. 1, pp. 227–248). Hillsdale, NJ: Erlbaum.
- Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 460–471.
- Smith, S. M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory & Cognition*, 10, 405–412.
- Spear, N. E. (1973). Retrieval of memory in animals. *Psychological Review*, 80, 163–194.
- Spear, N. E. (1978). *The processing of memories: Forgetting and retention*. Hillsdale, NJ: Erlbaum.
- Spear, N. E. (1981). Extending the domain of memory retrieval. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 341–378). Hillsdale, NJ: Erlbaum.
- Spear, N. E., Smith, G. J., Bryan, R., Gordon, W., Timmons, R., & Chiszar, D. (1980). Contextual influences on the interaction between conflicting memories in the rat. *Animal Learning & Behavior*, 8, 273–281.
- Speed, M. P. (1993). Muellierian mimicry and the psychology of predation. *Animal Behaviour*, 45, 571–580.
- Staddon, J. E. R. (1988). Learning as inference. In R. C. Bolles & M. D. Beecher (Eds.), *Evolution and learning* (pp. 59–78). Hillsdale, NJ: Erlbaum.
- Stewart, J., de Wit, H., & Eikelboom, R. (1984). Role of unconditioned and conditioned drug effects in the self-administration of opiates and stimulants. *Psychological Review*, 91, 251–268.
- Stolerman, I. (1992). Drugs of abuse: Behavioural principles, methods and terms. *Trends in Pharmacological Sciences*, 131, 170–176.
- Swartzentruber, D. (1991). Blocking between occasion setters and contextual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 17, 163–173.
- Swartzentruber, D., & Bouton, M. E. (1988). Transfer of positive contextual control across different conditioned stimuli. *Bulletin of the Psychonomic Society*, 26, 569–572.
- Thomas, D. R., McKelvie, A. R., Ranney, M., & Moye, T. B. (1981). Interference in pigeons' long-term memory viewed as a retrieval problem. *Animal Learning & Behavior*, 9, 581–586.
- Tulving, E. (1974). Cue-dependent forgetting. *American Scientist*, 62, 74–82.
- Turner, J. R. G. (1977). Butterfly mimicry: The genetical evolution of an adaptation. *Evolutionary Biology*, 10, 163–206.
- Underwood, B. J. (1948). "Spontaneous recovery" of verbal associations. *Journal of Experimental Psychology*, 38, 429–439.
- Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, 76, 559–573.
- Wagner, A. R. (1978). Expectancies and the priming of STM. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (pp. 177–210). Hillsdale, NJ: Erlbaum.
- Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale, NJ: Erlbaum.
- Wagner, A. R., & Brandon, S. E. (1989). Evolution of a structured connectionist model of Pavlovian conditioning (AESOP). In S. B. Klein & R. R. Mowrer (Eds.), *Contemporary learning theories: Pavlovian conditioning and the status of traditional learning theory* (pp. 149–190). Hillsdale, NJ: Erlbaum.
- Widrow, B., & Hoff, M. (1960). Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record*, 4, 96–194.
- Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, 2, 409–415.

Received February 2, 1994

Revision received February 16, 1994

Accepted February 16, 1994 ■