CHAPTER
18

Behavioral and Neural Correlates of Error Correction in Classical Conditioning and Human Category Learning

Mark A. Gluck
Rutgers University–Newark

To what extent are the processes of human learning analogous to the more elementary learning processes studied in animal-conditioning experiments? This question, and the broader goal of integrating mathematical models of animal and human learning, was the focus of my collaborative research at Stanford with Gordon Bower in the mid-1980s as well as my doctoral dissertation, which he supervised (Gluck & Bower, 1988a, 1988b, 1990). While working with Gordon, I also began a parallel line of research with another faculty member at Stanford, Richard Thompson. This research had the same conceptual starting point as the cognitive studies with Gordon, mathematical models of animal learning, but asked a different question: How are these learning principles embodied by neural circuits for various forms of classical conditioning (Gluck & Thompson, 1987; Thompson & Gluck, 1989, 1991).

In the late 1980s, these two research projects—one with Bower and the other with Thompson—shared only a common conceptual starting point. They were otherwise completely independent: The neuroscience research with Thompson made no direct links to cognition and the cognitive work with Bower made no direct links to neuroscience. These parallel projects continued throughout my graduate years at Stanford (1982–1987) as well as during several years of post-doctoral research at Stanford prior to my moving to Rutgers University–Newark in 1991. In the subsequent 15 years, my research has built upon the foundations of these two earlier research projects, extending them to create more direct bridges from neuroscience to human cognition. This newer cognitive neuroscience research fills in the gaps left by the earlier work with Bower and Thompson, showing how experimental and computational studies of the neural circuits for classical conditioning in animals has direct relevance for understanding the anatomy, physiology, neuropharmacology, and genetics of human learning, especially probabilistic category learning.

The remainder of this chapter is divided into four sections. In the first, I review the concept of error correction, and discuss how this learning principle has been a building block for models of both animal and human learning. Then, I turn to the neural substrates of error correction learning in classical conditioning, discussing the functional roles of three brain regions: the cerebellum, the basal ganglia, and the hippocampus. In the third section, I show how past bridges between animal and human learning (specifically my earlier doctoral dissertation research with Bower), provides a behavioral bridge for using models and data on the neural substrates of classical conditioning to inform our understanding of the cognitive neuroscience of human learning, especially probabilistic category learning. This research combines two methodologies, functional brain imaging and neuropsychological studies of patients with localized brain damage. In the fourth and final section, I briefly review the status of our understanding of the cognitive neuroscience of category learning, and some exciting new research directions that lie ahead.

ERROR CORRECTION IN LEARNING AND BEHAVIOR

For most of the first half of the 20th century, psychologists believed that as long as a cue (the conditioned stimulus, or CS) and an outcome (the unconditioned stimulus, or US) occurred closely together in time and nearby in space, an association would develop between them. However, in the late 1960s several psychological studies showed that pairing a CS and a US is not sufficient for conditioning to occur. Rather, for a CS to become associated with a US, it must provide valuable new information that helps an animal predict the future. Moreover, even if a given cue is predictive of a US, it may not become associated with that US if its usefulness has been preempted (“blocked”) by another co-occurring cue that has a longer history of predicting the US. For example, if a rat is first trained that a light predicts a shock, and later is trained that a compound stimulus of a light and tone together also predicts the shock, the rat will learn very little about the tone because the tone does not add any predictive information for the animal. This phenomenon, first described in animal conditioning by Leon Kamin, is known as blocking (Kamin, 1969). It demonstrates that classical conditioning occurs only when a cue is both a useful and a nonredundant predictor of the future.

The blocking effect challenged early theories of classical conditioning because it suggested that cues do not evoke conditioned responses based solely on their individual relationships with the US. Rather, blocking and other related experimental studies done in the late 1960s and early 1970s led to a new view of classical conditioning in which (a) cues that co-occur compete with each other to predict the US, and (b) a cue must impart reliable and nonredundant information about the expected occurrence of the US to produce effective conditioning. Apparently "simple" Pavlovian conditioning is not as simple as psychologists once thought it was. Even rats and rabbits act like sophisticated statisticians, sensitive to the relative informational value of cues in their environment.

Rescorla and Wagner's (1972) Error Correction Model of Classical Conditioning

In the early 1970s, two psychologists working at Yale University developed an elegant learning model to explain how animals might learn about the informational value of cues (Rescorla & Wagner, 1972). Rescorla and Wagner's key idea was that the changes in association between a CS and a US are driven by a prediction error, that is, the difference between whether or not the animal expects the US (i.e., the Expected US), and whether or not the US actually occurs (i.e., the Actual US). Rescorla and Wagner argued that if the occurrence of the US is unexpected, learning should occur proportional to the degree to which the US is surprising, where the surprise, that is the prediction error, is calculated as the difference between the Expected US and the Actual US. Key to their formulation was their assumption that an animal's expectation of the US is based on the sum of the strengths of all the CSs that are present on a trial. This allowed the model to account for many learning phenomena in which training to one cue can affect what is learned about other cues that are present in the same trials. In contrast, prior learning theories had assumed that each CS–US relationship is learned independently and were, thus, not able to address such cue–cue interactions during learning.

The Rescorla–Wagner model implied that a US that is totally unexpected given all the cues that are present (high error) should cause lots of learning whereas a US that is only partially expected (medium error) should result in less learning. The learning rule in the Rescorla–Wagner model is called an error correction rule because, over many trials of learning, it reduces—that is, "corrects"—the prediction error.

More than a quarter century after its publication, the Rescorla–Wagner model is generally acknowledged as the most powerful and influential formal model of learning ever produced by psychology. The model gained broad acceptance because it is simple, elegant, and explains a wide range of previously puzzling empirical results. It revealed an underlying order among a series of results that initially seem unrelated or even contradictory. The model also made novel and surprising predictions about how animals will behave in new experimental procedures, and experimenters rushed to test these predictions.

By virtue of its simplicity, the Rescorla–Wagner model does not account for all kinds of learning. Many researchers devoted themselves to showing how one or another addition to the model allows it to account for a wider range of phenomena—but with too many additions, the model loses some of its simplicity and appeal. Within its domain, the Rescorla–Wagner model combines explanatory power with mathematical simplicity. It takes an intuitively reasonable idea—classical conditioning is driven by a prediction error—and then pares away all but the most essential details, and uses this as a tool to explore implications of this idea that were not obvious before. The Rescorla–Wagner model is also a starting point from which many subsequent models were built, including the category-learning model of Gluck and Bower (1988) described next.

Gluck and Bower's (1988) Error Correction Model of Category Learning

To what extent are the processes of human learning analogous to the more elementary learning processes studied in animal-conditioning experiments? One consequence of the lack of communication between animal and human researchers in the 1960s is the fact that few, if any, animal researchers were aware that Gordon Bower and Tom Trabasso had demonstrated a form of blocking in human learning several years before the Kamin study (Trabbasso & Bower, 1964). During late 1960s and into the 1970s animal learning remained primarily concerned with elementary associative learning, whereas human-learning studies began to focus more on memory abilities, characterized in terms of information processing and rule-based symbol manipulation, approaches borrowed from the emerging field of artificial intelligence. Ironically, this historical schism between animal and human researchers occurred just as animal-learning theory was being invigorated by the Rescorla–Wagner model in the early 1970s.

Interest in relating human cognition to elementary associative learning was revived in the late 1980s by the expanding impact of computational "neural network" (or "connectionist") models of human learning. These models showed that many human abilities—including speech recognition, motor control, and category learning—emerge from configurations of elementary associations similar to those studied in conditioning paradigms (Rumelhart, McClelland, & the PDP Research Group, 1986).

One example of a connectionist model in cognition from that era is a simple neural network that Gordon Bower and I developed to model how people learn complex probabilistic categories (Gluck & Bower, 1988a). The study of category learning has been a central paradigm within cognitive psychology for more than 50 years. Category learning has aspects of both elementary associative learning as well as higher order cognition. On one hand, category learning can be viewed as a "cognitive skill" that shares many behavioral properties, and possibly some neural substrates, with motor-skill learning and conditioning. On the other hand, categorization underlies many higher order cognitive abilities. When a connoisseur distinguishes a cabernet from a merlot, or a doctor diagnoses a disease based on a pattern of symptoms, they are performing categorization. It is this dual nature—part elementary skill, part higher cognition—that makes category learning a valuable paradigm for studying fundamental aspects of human learning, at both the behavioral and neural levels of analysis.

The Gluck and Bower (1988) model of category learning was based on applying the Rescorla–Wagner model of animal conditioning to human learning. In our study, college students were asked to learn how to diagnose patients, according to which of two fictitious diseases they had, Midosis or Burlosis. The students reviewed medical records of fictitious patients, each of whom was suffering from one or more of the following symptoms: bloody nose, stomach cramps, puffy eyes, or discolored gums. During the study, subjects reviewed several hundred such medical charts, tried to diagnose each patient, and were then told the correct diagnosis. Initially, of course, the students had to guess; but with practice, they were able to diagnose the fictitious patients rather accurately. What helped them guess was that the different symptoms were differentially diagnostic of the two diseases. Thus, bloody noses were very common in Burlosis patients (but rare in Midosis) whereas discolored gums were common in Midosis patients (but rare in Burlosis). The other two symptoms, stomach cramps and puffy eyes, were only moderately diagnostic of either disease.

This kind of learning can be modeled using the network in Figure 18–1A. The four symptoms are represented by four input nodes at the bottom of the network and the two diseases correspond to the two output nodes at the top of the network. The weights between the symptoms and the diseases are updated according to the learning rule from the Rescorla–Wagner model, much as if the symptoms were CSs and the diseases were alternate USs.

Learning and performance in the model works as follows: If on a particular trial, the symptoms "Bloody Nose" and "Stomach Cramp" are presented, then this is modeled by turning "on" the corresponding input nodes (stippling in Fig. 18–1). These act like two CSs present on a conditioning trial. In contrast to the classical conditioning paradigms described earlier, where there is one US (e.g., a shock), here there are two possible outcomes, the diseases Burlosis and Midosis. For each outcome category, there is a teaching node that provides error-correcting feedback with the correct (actual) category for each input training pattern. In Figure 18–1, the correct category is Burlosis. Thus, activating two features

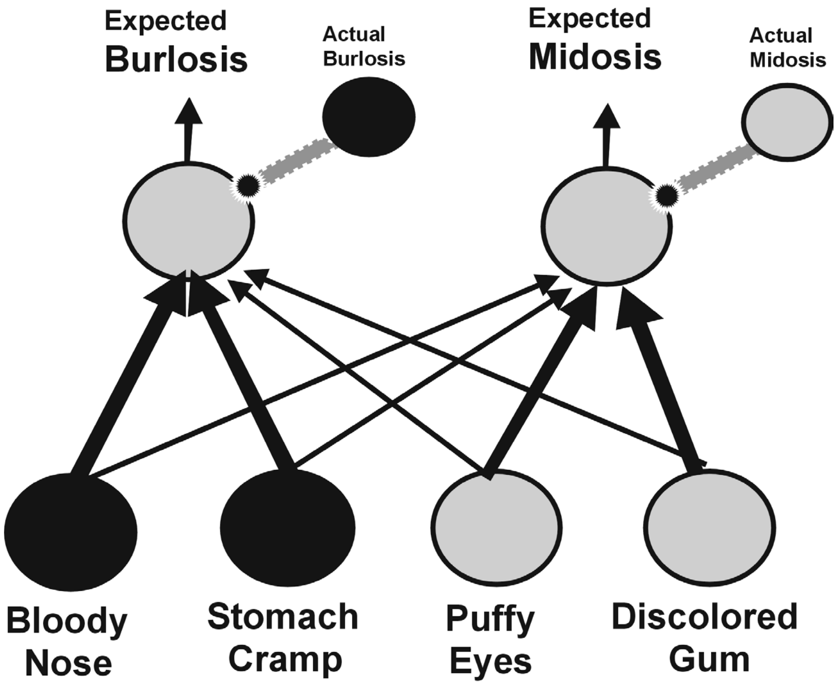


Figure 18–1A. Medical diagnosis network applying Rescorla–Wagner model to human category learning (after Gluck & Bower, 1988a). The Model. The weights from bloody nose to Burlosis and from discolored gums to Midosis are thick indicating highly diagnostic relationships. The other cues are of only moderate diagnosticity. The input nodes for bloody nose and stomach cramp are shown activated by the dark fill. For each outcome category there is a teaching node that provides error-correcting feedback with the correct (actual) category for each input training pattern. In this case, the correct category is Burlosis.

(two input nodes) causes activity to travel up four weights, two to Burlosis and two to Midosis as shown in Figure 18–1.

By analogy with the Rescorla–Wagner model, the output node activations are equivalent to the network's expectation of one disease versus another, and the correct answer (the disease name given by the experimenter) was then used to modify the weights to reduce the error between the expected disease and the actual disease category outcome, according to Rescorla–Wagner's error correction learning rule. The network model shown in Figure 18–1A incorporates nothing more than the learning principle of the 1972 Rescorla–Wagner conditioning model. Nevertheless, this "animal-conditioning" model of human cognition accounts for many aspects of how people in our experiments classified different patients.

With four possible symptoms, there are 16 possible patient charts that can be constructed depending on whether each of the four symptoms is present or absent. We actually used only 14 of these, eliminating the charts with no symptoms (all absent) or all four symptoms (all present). After subjects had completed a long series of training trials, we asked if the model could predict the proportion of times that each of the 14 patterns was classified as Burlosis versus Midosis by their subjects. To generate this prediction, we looked at two output nodes, *Expected-Burlosis* and *Expected-Midosis,* for each of the 14 patterns. If, for a particular symptom pattern, such as "Bloody Nose & Stomach Cramp," the output values were *Expected-Burlosis* = 80 and *Expected-Midosis* = 20, then, we argued, the subjects should likely classify this pattern as Burlosis 80% of the time and as Midosis 20% of the time. In this way, we calculated a predicted proportion of "Burlosis" responses for each of these 14 patterns based on their model and compared this to the actual proportion of subjects who responded "Burlosis" to these patterns during the final 50 trials of the experiments (Gluck & Bower, 1988a).

The results of this analysis are shown in Figure 18–1B, where each of the 14 patterns is represented by a dot. The location of each dot corresponds (on the horizontal axis) to the model's predicted proportion (ranging from 0 to 1), whereas its location on the vertical axis corresponds to the actual experimental data. Thus, the "Bloody Nose & Stomach Cramp" patient from Figure 18–1 who has a predicted proportion of 80% Burlosis categorization would be located as a dot at 0.8 on the horizontal axis. If, indeed, the subjects in this experiment did label this pattern as Burlosis on 80% of the trials, then the dot for "Bloody Nose & Stomach Cramp" would be found at the point (0.8,0.8) in this graph. Thus, the better the fit of the model the more likely that each of the 14 patterns (dots) would lie on the straight line from (0,0) through (1,1). As you can see from Figure 18–1A, the fit is excellent.

In addition to these fits, the model was applied to many other types of data from these and other experiments. It was able to account for the relative differences in difficulty among many different tasks (i.e., why some tasks are harder than others and take longer to solve) and it predicted some surprising generalization behaviors when people, following category learning, were later asked to predict the marginal probabilities of different categories given the presence of individual features (Gluck & Bower, 1988a, 1988b). For learning more complex discrimination rules in which sensitivity to the relationships between stimulus features was necessary, we borrowed again from Rescorla and Wagner, adopting their convention of including configural nodes that represented the unique configuration of various pairs of features (e.g., bloody nose and stomach cramps both being present); again, this approach showed a remarkable ability to explain a wide range of human category-learning behaviors (Gluck & Bower, 1988b; Gluck, Bower, & Hee, 1989).

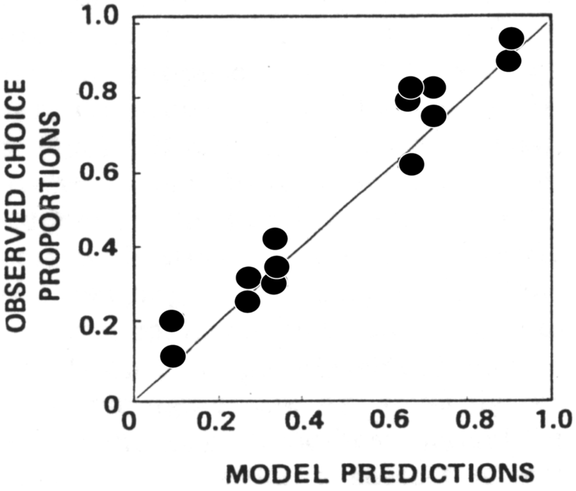


Figure 18–1B. Fits of the Rescorla–Wagner network model for pattern classification of 14 items (from Gluck & Bower, 1988a). Each pattern is represented by a dot whose location is determined by the model predictions (x-axis) and the actual pattern classification proportions (y-axis). The fact that the 14 dots lie very close to the diagonal line indicates a very close fit of model to data.

THE NEURAL SUBSTRATES OF ERROR CORRECTION LEARNING IN CLASSICAL CONDITIONING

The behavioral studies described in the previous section demonstrate that error correction learning is common to both animal conditioning and human category learning. But how is error correction computed in the brain? In this section, I briefly review what is known about the neural substrates of error correction in animal studies of classical conditioning, including the work of Richard Thompson and myself on the cerebellum and aversive conditioning of motor reflexes, as well as the work of Wolfram Shultz and colleagues on the basal ganglia and midbrain dopamine neurons and their role in appetitive conditioning. This leads into discussing the role of the hippocampus in modulating both forms of learning. The following section then builds on this discussion to address what is known so far about the neural substrates of error correction in human learning.

THE CEREBELLUM AND ERROR CORRECTION IN AVERSIVE CONDITIONING OF MOTOR REFLEXES

In the early 1980s, Richard Thompson and his coworkers discovered that small lesions in the cerebellum of rabbits permanently and completely prevented the acquisition of new classically conditioned eyeblink responses and abolished retention of previously learned responses (Thompson, 1986). As shown in Figure 18–2, the cerebellum has two main layers. The top surface of the cerebellum is the cerebellar cortex (which includes the Purkinje cells). Below the cerebellar cortex lies the interpositus nucleus, one of the cerebellar deep nuclei.

To follow the pathways in and out of the cerebellum, we begin with the CSs that project first to an area in the brain stem called the pontine nuclei. The pontine nuclei include different subregions for each kind of sensory stimulation. Thus, a tone CS would project to one area of the pontine nuclei and a light CS to another. This CS information then travels up to the deep nuclei of the cerebellum along mossy fibers, which branch in two directions. First, they make contact in the deep nuclei with the interpositus nucleus. Second, they project up to the cerebellar cortex (via a few other cells not shown) and then connect to the Purkinje cells in the cerebellar cortex. The second sensory input pathway is the US pathway. An air puff to the eye, the US, activates the inferior olive, a structure that activates the interpositus in the deep nucleus of the cerebellum. In addition, a second pathway from the inferior olive projects up to the cerebellar cortex via climb-

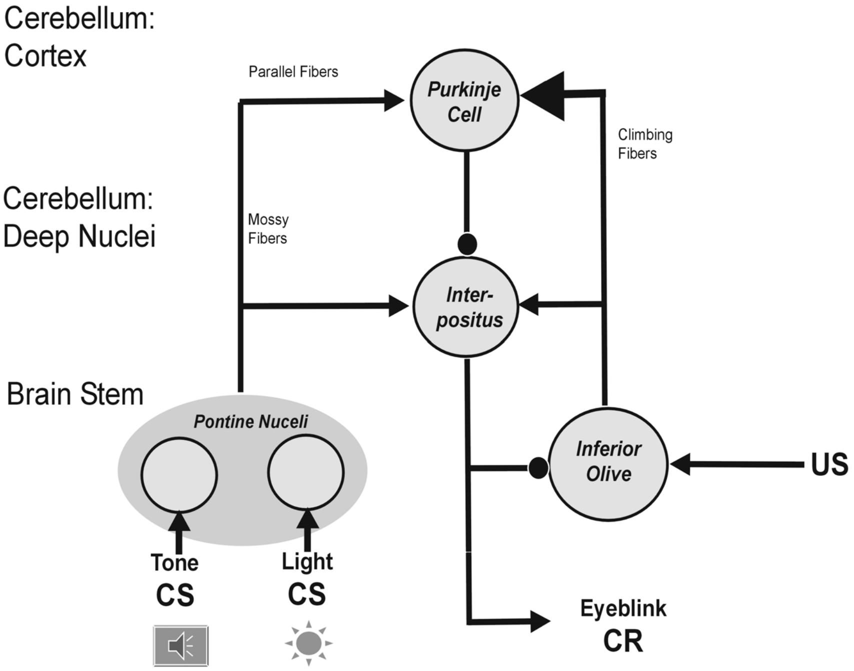


Figure 18–2. Cerebellar circuits showing the CS pathway, the US pathway, and the CR pathway, projecting up from the sensory cues into the cerebellar cortex and deep nuclei. Excitatory synapses are shown as arrows and inhibitory synapses terminate with a filled circle.

ing fibers. Complementing these two input pathways is a single-output pathway for the CR (conditioned response), which originates with the Purkinje cells. The Purkinje cells project down from the cerebellar cortex into the deep nuclei where they form an inhibitory synapse with the interpositus. The interpositus is the only output from the cerebellum; activity in the interpositus projects to the motor cortex which, in turn, projects to the muscles in the eye to generate the eyeblink CR.

There are two sites in the cerebellum where CS and US information converge and, thus, where information about the CS–US association might be stored: (a) the Purkinje cells in the cerebellar cortex and (b) the interpositus nucleus. These two sites of convergence are intimately interconnected through an output pathway; the Purkinje cells project down to the interpositus nucleus with strong inhibitory synapses, as shown in Figure 18–2.

Note that there is also an additional pathway within the cerebellum that we have not yet discussed. This inhibitory feedback pathway projects from the interpositus nucleus to the inferior olive. Thus, in a well-trained animal that makes a CR and activates the interpositus nucleus, this activity will, in turn, inhibit the inferior olive carrying US information (Sears & Steinmetz, 1991). Thus, activity in the inferior olive will reflect the *Actual-US* less (due to inhibition) the *Expected-US,* where the *Expected-US* is measured by the interpositus activity, which drives the CR. *Actual-US—Expected-US.* Note that this is the same difference (*Actual-US—Expected-US*) that the Rescorla–Wagner model uses to calculate the error on a trial, and which is then used to determine how much learning should accrue to the CS association weights. In several papers, Richard Thompson and I developed computational models that showed how these circuits could implement the essential error correction principle of the Rescorla–Wagner model, along with various other aspects of timing and response behaviors (Donegan, Gluck, & Thompson, 1989; Gluck, Allen, Myers, & Thompson, 2001; Thompson & Gluck, 1991).

Our interpretation for how the cerebellum computes the Rescorla–Wagner model's error correction procedure implies that Kamin's blocking effect (the clearest experimental evidence for error correction learning) should depend on that inhibitory pathway from the interpositus to the inferior olive. This prediction was experimentally confirmed in a later study by Thompson and colleagues, who disabled this inhibitory pathway and, in doing so, showed that they could *block* blocking (Kim, Krupa, & Thompson, 1998). More generally, our computational modeling, along with various other experimental studies, argues that the cerebellum acts as a predictive system that learns through error correction principles to make anticipatory adjustments in timing-sensitive behaviors.

The Basal Ganglia and Error Correction in Appetitive Conditioning

The previous section showed that the inferior olive in the cerebellum codes for the prediction error during eyeblink conditioning, much as described by the Rescorla–Wagner model. The inferior olive activity is high when the air puff US is unexpected, drops down to baseline when the US is predicted, and shows below-baseline firing rates when an expected US does not occur (i.e., when the error term is negative). These cerebellar circuits, however, are not responsible for all forms of classical conditioning, but only for conditioning of discrete well-timed motor reflexes like the eyeblink response. What about other forms of classical conditioning, especially those where the US is a positive reward, such as food or drink?

A series of electrophysiological recording studies in monkeys led researchers to suggest that dopamine neurons in the midbrain play a critical role in reward-related learning (for a review, see Schultz, 1998; Schultz, Dayan, & Montague, 1997). Specifically, these dopamine neurons respond with strong bursts of activity to unexpected rewards (but not to expected rewards), and show a decrease in firing when an expected reward fails to occur. Thus, these dopamine neurons appear to behave in appetitive conditioning (where the US is a positive rewarding stimulus) very much like the inferior olive cells do during motor-reflex conditioning to an aversive US: They code for the prediction error. More generally, work by Schultz and others has confirmed that dopamine neurons in the midbrain (both in substantia nigra compacta and in the ventral tegmental area) play a role in implementing the error-correcting principles of the Rescorla–Wagner model in certain appetitive forms of classical conditioning, in ways broadly analogous to the role of the cerebellum in aversive conditioning of motor-reflex responses.

What Does the Hippocampus Do in Classical Conditioning?

If the cerebellum is essential for aversive conditioning of well-timed motor reflexes and midbrain dopamine neurons are key for conditioning of appetitive reward prediction tasks, what, if any, role does the hippocampus play in these forms of classical conditioning? For half a century, it has been appreciated that the hippocampal region plays a critical role in acquisition of new memories, particularly rapidly acquired memories for autobiographical events, sometimes collectively called episodic memory (e.g., Squire, 1987). More recently, data from human and animal studies have documented that the hippocampal region is also involved in many kinds of incrementally acquired learning, including simple associative learning such as conditioning and category learning. What does the hippocampus contribute to classical conditioning, above and beyond the functions subserved by the cerebellum and basal ganglia?

After moving to Rutgers–Newark in 1991, I began a new program of hippocampal modeling with my (then) postdoctoral fellow, Catherine Myers. Together, we developed a neural network model of cortico-hippocampal processing to account for data from studies of classical conditioning in animals with lesions to their hippocampal region (Gluck & Myers, 1993, 2001; Myers & Gluck, 1994). The model conceptualizes the brain as a series of interacting modules, each implementing the information-processing function subserved by a particular brain region, but without regard for whether that function is implemented in a biologically plausible way.

As described earlier, the cerebellum is the substrate for storage and expression of learned CS–US associations in motor-reflex conditioning (Thompson, 1986). We adapted our earlier Thompson–Gluck cerebellar model of Figure 18–2 into a simpler connectionist network model shown in the left of Figure 18–3 (Gluck, Myers, & Thompson, 1994). This network learns to map from inputs specifying the presence of CSs and contextual cues, to a pattern of activation in an internal layer of nodes via a layer of weighted connections. This internal activation pattern constitutes a remapping or rerepresentation of the input, which is then mapped to output driving the behavioral CR via a second layer of weighted connections. On each trial, the system "error" is the difference between the actual response (CR) and the desired response (US). An error correction learning rule (analogous to the Rescorla–Wagner model) was used to modify the weights between the internal-layer and output-layer nodes, proportional to this error. However, no such error measure is defined for the internal-layer nodes, and so this error correction rule cannot be used to modify the lower layer of weights. As a result, no learning takes place in the lower layer and thus, the "internal representation" of stimuli at the intermediate layer of nodes is fixed if the hippocampal region model is missing (i.e., lesioned). Nevertheless, for many simple problems,

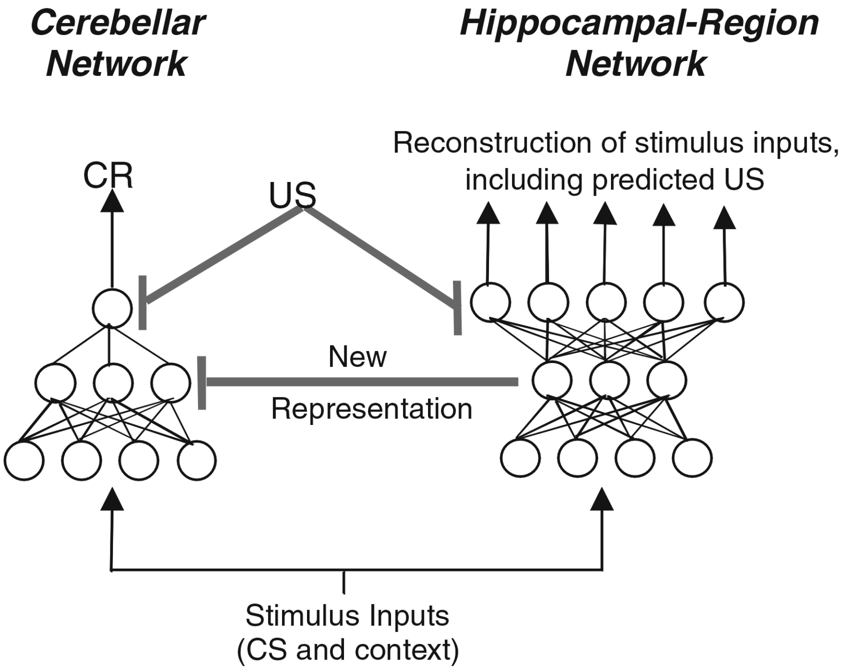

Cerebellar Network
Hippocampal-Region Network
Reconstruction of stimulus inputs, including predicted US
CR
US
New Representation
Stimulus Inputs (CS and context)

Figure 18–3. The Gluck and Myers (1993) cortico-hippocampal model. The hippocampal region forms new representations that compress redundancy while differentiating predictive information; these new representations can be adopted by long-term memory sites such as the cerebellum.

this system can learn appropriate CS–US association and produce a behavioral CR similar to empirical learning curves (Gluck et al., 1994). The abstract connectionist model of the cerebellar contributions to classical conditioning, shown on the left of Figure 18–3, can be directly related to the same information-processing capabilities of the more physiologically detailed and biologically realistic model in Figure 18–2: Both alter CS–US associations according to the error-correcting principle of the Rescorla–Wagner model.

With this simplified cerebellar model of conditioning, we were now able to ask: What does the hippocampus do? Catherine Myers and I proposed that the hippocampal region contributes to this cerebellar learning by developing new representations that encode stimulus–stimulus regularities (Gluck & Myers, 1993). In particular, if two stimuli reliably co-occur or are otherwise redundant, their representations become compressed, or more similar. Conversely, if two stimuli predict different future events, their representations become differentiated, or less similar.

We implemented this theory in a connectionist network model as shown in the full network model of Figure 18–3 (Gluck & Myers, 1993, 2001). Hippocampal-region processing is implemented via a network that learns to map CS inputs, through an internal node layer, to outputs that reconstruct those inputs and also predict the US. This network, unlike the cerebellar network, is able to modify both layers of weighted connections through a learning algorithm such as error back-propagation (Rumelhart et al., 1986). In the process, internal-layer nodes form a representation of the input that tends to compress redundant information while differentiating information that predicts the US, just as required by our theory.

A random recoding of the hippocampal-region network's internal-layer activations becomes the "desired output" for each node in the internal layer of the cerebellar network, and each node's error is the difference between this desired output and its own actual output. The cerebellar network then uses the error-correcting rule to adapt its lower-layer weights, just as it uses simple error correction learning to adapt its upper-layer weights. Over time, representations develop in the internal-layer nodes of the cerebellar network that are linear recombinations of the new representations developed by the hippocampal region network. Within this model framework, broad hippocampal-region damage is simulated by disabling the hippocampal-region network, leading to a network model in which the cerebellum alone processes information without modulating input from the hippocampal region. In this lesioned model, the error-correcting cerebellar network cannot adopt any new representations, although it can still learn to map from its existing representations to new behavioral responses by modifying its upper layer of weights.

Our model of hippocampal-region function correctly accounted for data showing that hippocampal-region damage does not impair simple delay conditioning but does impair more complex behaviors including contextual sensitivity and sensory preconditioning (Gluck & Myers, 1993, 2001; Myers & Gluck, 1994). It also made several novel predictions. For example, it predicted that learned irrelevance (slower CS–US learning following uncorrelated CS–US exposure) should be disrupted following hippocampal-region damage; we confirmed this in our lab at Rutgers in studies with rabbit eyeblink conditioning (Allen, Chelius, & Gluck,

2002) as well as humans (Myers et al., 2000). Similarly, our model expected that acquired equivalence (transfer of associations between objects previously associated with similar consequences) should be disrupted following hippocampal-region damage; this has also been confirmed in animals (Coutureau et al., 2002) and in studies we did in humans (Myers et al., 2003). In all these tasks, the common theme, as predicted by the model, is that the hippocampal region is not required for simple stimulus–response learning, but is required for learning about contextual or stimulus–stimulus regularities that support subsequent transfer generalization, phenomena that are not predicted by the error correction learning principle of the Rescorla–Wagner model.

As noted previously, many behavioral phenomena that cannot be explained by the Rescorla–Wagner model are not found in animals that have lesions to the hippocampal region. This suggests that the Rescorla–Wagner model may be better described as a model of the cerebellar contributions to motor-reflex conditioning in hippocampal-lesioned animals than as a model of conditioning in healthy, intact animals. Thus, the limitations of the Rescorla–Wagner model might now be reinterpreted as symptoms that this mathematical model of learning from the 1970s isn't really dead, just "brain damaged." That is to say, the model applies to the brain regions responsible for error correction learning such as the cerebellum, but does not explain the additional contributions of the hippocampal region.

Within its limited domain, the early Gluck and Myers model was reasonably successful at providing an account of the role of the hippocampal region in associative learning. However, it was implemented without particular regard for the anatomical or physiological details of the brain substrate. In part, this reflected the state of the empirical literature at the time: Most data on hippocampal-region function were based on lesion studies using techniques like ablation that were not sufficiently selective to allow complete destruction of a specific brain structure without conjoint damage to other nearby structures and to fibers of passage. Newer lesion techniques (such as neurotoxic lesions using ibotenic acid; see also Jarrard, 2002) have since allowed the accumulation of a large body of data contrasting, for example, the selective effects of entorhinal versus hippocampal damage, and electrophysiological recording studies have provided additional insights and constraints. As a result, there is now a sufficient body of empirical data to constrain a model differentiating these structures; this is the focus of current modeling efforts at Rutgers, including collaborative work with a former postdoctoral fellow in my lab, Martijn Meeter, who is now at the Free University in Amsterdam.

THE COGNITIVE NEUROSCIENCE OF ERROR CORRECTION IN PROBABILISTIC CATEGORY LEARNING

To summarize the results discussed so far: The cerebellum and basal ganglia can be understood as implementing the error correction mechanisms for learning described by the Rescorla–Wagner model for two forms of classical conditioning whereas the hippocampus operates during all forms of conditioning to create novel stimulus representations that reflect stimulus–stimulus regularities in the environment. What do these results imply about the neural bases of human learning? One avenue for seeking linkages between animal conditioning and human learning is to look at studies of classical conditioning in humans. Indeed, there exists an extensive literature on classical conditioning of the human eyeblink response, including those that used functional brain imaging in healthy normal people and studies of behavior in clinical populations. The conclusion that can be drawn from this is that the neural substrates for motor-reflex conditioning appear to be identical in humans and other animals (Daum et al., 1993; Gabrieli et al., 1995; Logan & Grafton, 1995).

Another avenue for seeking linkages between animal research and human learning is by using more cognitive tasks that employ analogous error correction principles of learning. This is where the earlier work that Gordon and I did in the late 1980s becomes relevant once again. Given our prior results showing that people learn probabilistic categories using *behavioral* principles of error correction analogous to those seen in classical conditioning, we can now ask: Are there also analogous *neural* mechanisms involved in human category learning that are similar or identical to those involved in classical conditioning? This leads to two specific questions: First, in human category learning, what are the neural mechanisms for error correction based on cognitive feedback? Second, does the hippocampal region play an analogous role in category learning as it does in classical conditioning creating novel stimulus representations? These two questions have driven my lab's more recent research on the cognitive neuroscience of category learning.

Probabilistic Category Learning

Beginning with the Gluck and Bower (1988a) studies reviewed earlier, category-learning research in my lab has focused primarily on learning probabilistic categories. These are categories in which there is no clear-cut rule for membership. Rather, various features are more, or less, probabilistically associated with one category or another. For example, "red sky at night" is a feature that is partially correlated with the category of "good weather tomorrow" but this feature is not a perfect rule for predicting the weather—only a useful heuristic. The weather might be better predicted, on average, by employing evidence from several such features, although even then, it might be impossible to predict the upcoming weather with 100% accuracy.

In the mid-1990s, we developed at Rutgers several novel probabilistic category-learning tasks based on variations of the earlier studies by Gluck and Bower (Gluck & Bower, 1988a, 1988b). The most well-known—and widely adopted—of our new category-learning tasks is often referred to as the "weather prediction" task (Gluck, Shohamy, & Myers, 2002; Knowlton, Squire, & Gluck, 1994; Poldrack et al., 2001). It uses four cards with geometric patterns as stimulus features, as shown in Figure 18–4A. On each trial, a subject sees one or more of these cards and is asked to predict whether the next day's weather will be rain or sunshine, as illustrated in Figure 18–4B.

The actual weather outcome is determined by a probabilistic rule based on the cards: Each card predicts rain or sunshine with a fixed probability as shown in Figure 18–4A, based on the same categories used in Gluck and Bower (1988a). Thus, the card with squares (S1) is strongly predictive of rain whereas the card with triangles (S4) is strongly predictive of sunshine. The other two cards have more intermediate statistical relationship with the two outcome categories. The actual outcome is based on the cumulative probabilities associated with all cards present on a trial. The probabilistic relationships between cues and outcomes ensures that it is impossible for subjects to learn the categorization with complete certainty, although it is possible to achieve significant learning by inducing how diagnostic each card is for each category.

In an early study of this task, we collaborated with Larry Squire and Barbara Knowlton using amnesic patients from the San Diego area who presented with a variety of etiologies including those with Korsakoff's syndrome, unknown lesions, as well as more focal medial temporal lobe damage. These amnesic patients learned the weather prediction task at about the same rate as control subjects, improving from chance performance (50% correct) to approximately 65% correct over the first 50 trials (Knowlton et al., 1994). With extended training, however, control subjects outperformed amnesic patients. In a more recent study, however, we used a group of patients with more localized hippocampal-region damage all of whom had a common etiology for their amnesia: hypoxia, the loss of oxygen to the brain (Hopkins et al., 2004). We found that these amnesics were uniformly impaired at two forms of probabilistic category learning, both early and late in training, in

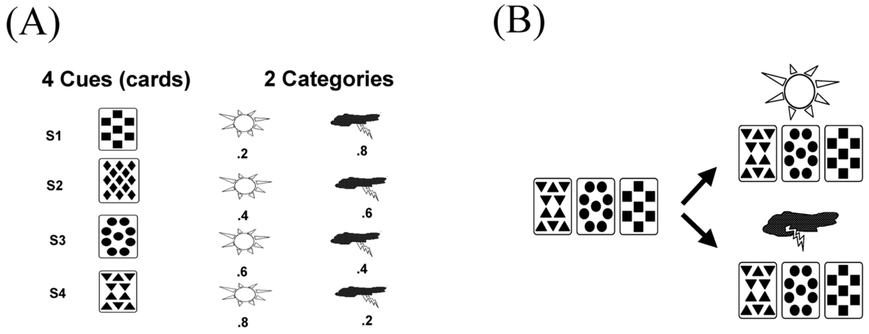


Figure 18–4. (a) Four cards with geometric patterns are each related probabilistically to two different outcome categories, good weather and bad weather. (b) On each trial, a subject sees one or more of these cards (shown on the left) and is asked to predict whether the next day's weather will be sunshine (top right) or rain (bottom right).

contrast to the previous report by Knowlton et al., which had found deficits only later in training.

Together these results suggest that the weather prediction task, using the category structures from Knowlton et al. (1994) and Gluck and Bower (1988a), requires considerable hippocampal mediation. The Gluck and Myers (1993) model would argue that this reflects the many stimulus–stimulus relationships that can be used by an intact hippocampal region to support learning in this category structure, even if the task, being linearly separable, could, in principle, be learned without recourse to configural cues. This does not, however, imply that all forms of category learning, probabilistic or otherwise, will depend on an intact hippocampal region. Rather, simpler forms of category learning without significant stimulus–stimulus correlations are expected by our model to be largely independent of the hippocampal region. Consistent with this prediction, we have more recently shown that hypoxic amnesic patients are able to acquire a simpler category learning tasks at about the same rate as healthy controls (Shohamy, Myers, Geghman, Sage, & Gluck, 2006).

Functional Brain Imaging of Probabilistic Category Learning

Our prior computational models of hippocampal-region function (Gluck & Myers, 1993, 2001) suggest that hippocampal region develops new stimulus representations that are eventually acquired by other brain regions. In the last few years, support for the applicability of our model of hippocampal-region function to human learning has come from studies using functional brain imaging. Our model predicts that the hippocampal region should be very active early in category learning tasks when participants are learning about stimulus–stimulus regularities and evolving new stimulus representations, but that the hippocampal region should be less active later in training when other brain regions (e.g., cerebellum or basal ganglia) are using these representations to perform the behavioral response. In collaboration with Russ Poldrack at UCLA, we conducted a functional neuroimaging (fMRI) study of normal humans learning the weather prediction task. As expected by our model, we found that activity in the hippocampal region was high early in training and then tapered off; in contrast, basal ganglia activity was low at first and increased during training (Poldrack et al., 2001).

In Poldrack et al. (2001), we also examined whether activity in the basal ganglia and MTL was modulated by task demands. In particular, we compared two versions of the weather prediction task: the standard feedback-based version of the task, and an observational learning version in which subjects simply viewed stimulus–outcome pairs on each trial, and were later tested on these associations. Although learning on these two versions was equivalent in terms of percent optimal responding during a final testing phase, basal ganglia activation and MTL deactivation were significantly stronger during the feedback-based version of the task compared to the observational version of the task. This is consistent with the view that the basal ganglia (but not the MTL) are key for learning based on error-correcting feedback, but not during observational learning in which no response and no feedback is involved.

More recently, in collaboration with Daniel Weinberger and colleagues at NIMH, we have shown that probabilistic category learning engages neural circuitry that includes both the prefrontal cortex and the caudate nucleus of the basal ganglia, two regions that show prominent changes with normal aging (Fera et al., 2005). When trained on the weather prediction task, young and older adults displayed equivalent learning curves, used similar strategies, and activated analogous brain regions as seen using fMRI. However, the extent of caudate and prefrontal activation was less, and parietal activation was greater, in older participants. This suggests that some brain regions, such as the parietal cortices, may provide compensatory mechanisms for healthy older adults in the context of deficient prefrontal cortex and caudate nuclei responses. Further research will be required to better understand these age-related changes, but the initial study points to the promise of using probabilistic category learning tasks as a means to understand changes in neural function over the life span.

The Basal Ganglia and Category Learning in Parkinson's Disease

If the basal ganglia are key for error correction learning that is based on feedback, we should expect people with damage to the basal ganglia to show impairments on probabilistic category learning. One such population is patients with Parkinson's disease (PD) who have a profound loss of dopamine containing neurons in the substantia nigra pars compacta (SNc), leading to dopamine depletion in the basal ganglia. These are among the dopamine cells that Wolfram Schultz and colleagues have identified with error correction computations in reward prediction conditioning, as described earlier.

The loss of dopamine in PD leads most prominently to a loss of motor control. However, recent studies have shown that the loss of dopamine that occurs in PD also leads to a variety of cognitive deficits, especially tasks that involve incremental learning of associations between cues and outcomes based on error-correcting feedback (Knowlton, Mangels, & Squire, 1996; Myers, Shohamy et al., 2003a, 2003b; Shohamy, Myers, Grossman, et al., 2004; Shohamy, Myers, Onlaor, & Gluck, 2004). These findings suggest that midbrain dopamine may be particularly important for learning that involves the incremental acquisition of stimulus–outcome associations via error-correcting feedback. This is consistent with converging evidence from the functional imaging studies described earlier (Poldrack et al., 2001).

To explore this issue further, Daphna Shohamy, who was then a graduate student in my laboratory, initiated a series of studies of category learning in Parkinson's patients. In the first study, we looked at how Parkinson's patients learn the weather prediction task over 3 days of extensive training. As shown in Figure 18–5, the

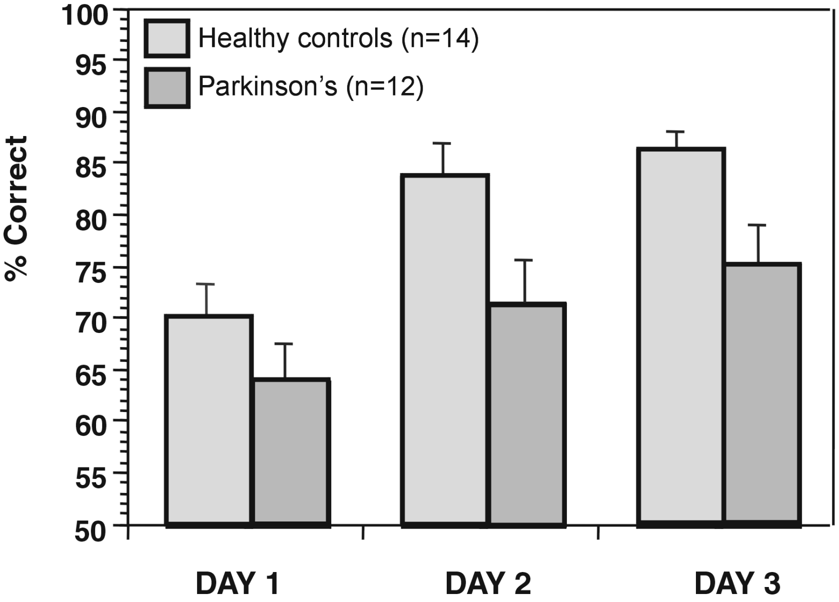


Figure 18–5. Parkinson's patients are slower to learn the weather prediction task over three days of training. Data from Shohamy, Myers, Onlaor, and Gluck (2004).

patients were significantly impaired relative to matched controls over the course of learning (Shohamy, Myers, Onlaor, & Gluck, 2004). Additional analyses showed differences in the learning strategies used by these two groups. Healthy controls all began to solve the task by using single features and then shifted over the course of 3 days of training to using more complex rules that integrated information from multiple cues. In contrast, the Parkinson's patients continued throughout the study to use the less accurate simpler single-cue rules.

More recently, Shohamy, Myers, and I have followed up on this study with further analyses of probabilistic category learning, using a variant on the weather prediction task in which the stimuli are digital photographs of Mr. Potato Head figures that have one or more facial features (e.g., moustache, hat, glasses, and bowtie) that can each be present or absent (Fig. 18–6). Rather than predicting the weather, the subjects are asked to predict which flavor of ice cream (vanilla or chocolate) Mr. Potato Head prefers. Each facial feature is probabilistically associated with each outcome, just as in the weather prediction experiment.

In several recent studies , we have used this as a cover story for probabilistic category learning, using category structures analogous to those in the weather prediction task and the earlier Gluck and Bower studies. We found that subjects find the Mr. Potato Head task more engaging and appealing. It also allows for a wider range of possible features and task demands, because of the large number of facial features available.

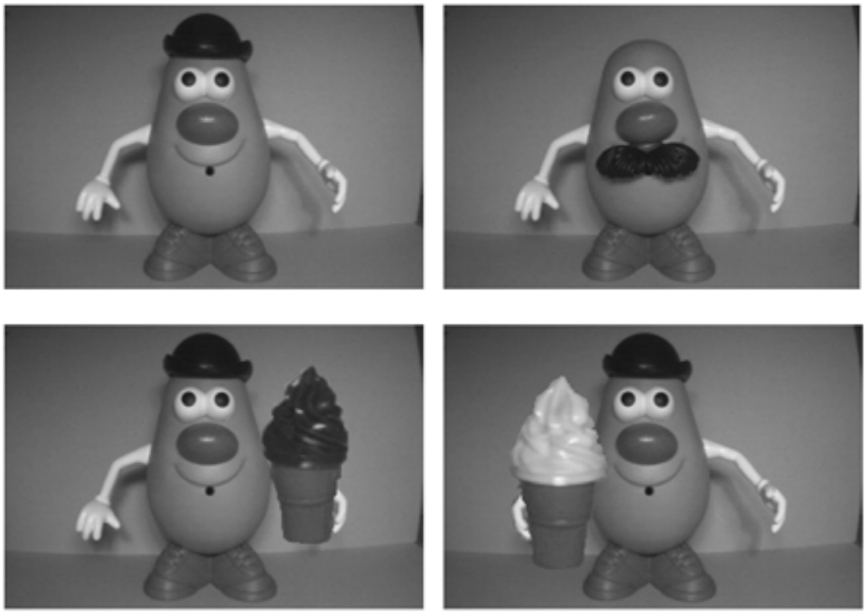
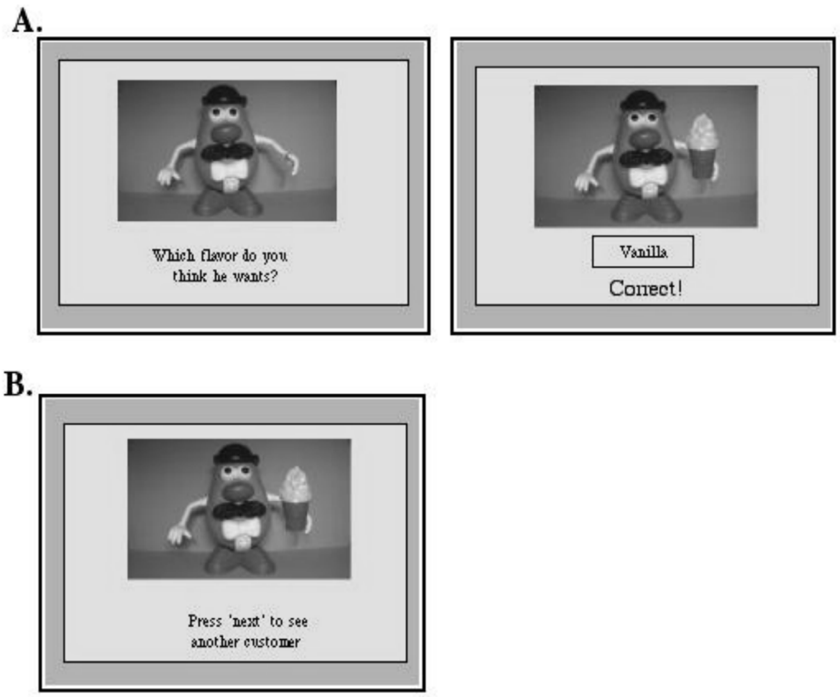


Figure 18–6. Examples of stimuli used in Shohamy, Myers, Grossman, et al. (2004) including Mr. Potato Head figures with different features (hat and moustache) and different category membership feedback (vanilla and chocolate ice cream).

Using this Mr. Potato Head task, we began a series of studies to examine how manipulations of training procedures would affect learning performance in Parkinson's patients (Shohamy, Myers, Grossman, et al., 2004). In one study, we trained participants using standard "feedback" training. On each trial, subjects saw the stimulus, responded with a guess as to the outcome, and then received feedback as to whether that response was correct (Fig. 18–7A). In a second, "observational" version, subjects saw the same stimuli, but are shown the correct outcome (Fig. 18–7B). To assess learning in the "observational" group, subjects were then given test trials in which they see the stimulus and must respond with the outcome information—but no feedback is provided. Thus, we can compare learning under observation or feedback by comparing the last block of the feedback training (last 50 trials) with the 50 test trials following observation training.

Based on the proposed role of the dopamine signals in reward feedback processing by the basal ganglia and the aforementioned imaging data with Poldrack, we predicted that PD patients would be impaired at the feedback-based version (during both training and transfer testing) but would perform as well as controls on the observational version on the transfer test phase.

As shown in Figure 18–8, performance was impaired in the PD patients who had been trained in the feedback condition, but not in those trained in the obser-



Which flavor do you think he wants?

Correct!

Press 'next' to see another customer

Figure 18–7. (a) Standard feedback training in which subjects first see the stimulus alone, make a categorization response (*vanilla* or *chocolate*), and then get feedback. (b) Observational training in which subjects are exposed to the stimuli with their correct category and are not required to make a categorization response, nor get any feedback.

vational condition. Thus, as predicted by our hypothesis, the PD patients are impaired at learning that involves incremental feedback, but are not impaired at learning cue–outcome associations if those are presented in a nonfeedback manner.

These results provide behavioral evidence that the basal ganglia are necessary for feedback-based learning in a cognitive task. The results provide a direct confirmation of a prediction inspired by our previous neuroimaging results with healthy humans (Poldrack et al., 2001), which had demonstrated differences in engagement of basal ganglia and midbrain dopaminergic regions between feedback-based and observational learning.

More recent studies in our laboratory, using novel forced-choice and concurrent discrimination tasks developed by Catherine Myers, have also demonstrated a double dissociation between MTL and basal ganglia contributions to learning within single associative-learning tasks (Myers, Shohamy, et al., 2003; Shohamy et al., 2006). We found that PD patients were slow to acquire a discrimination

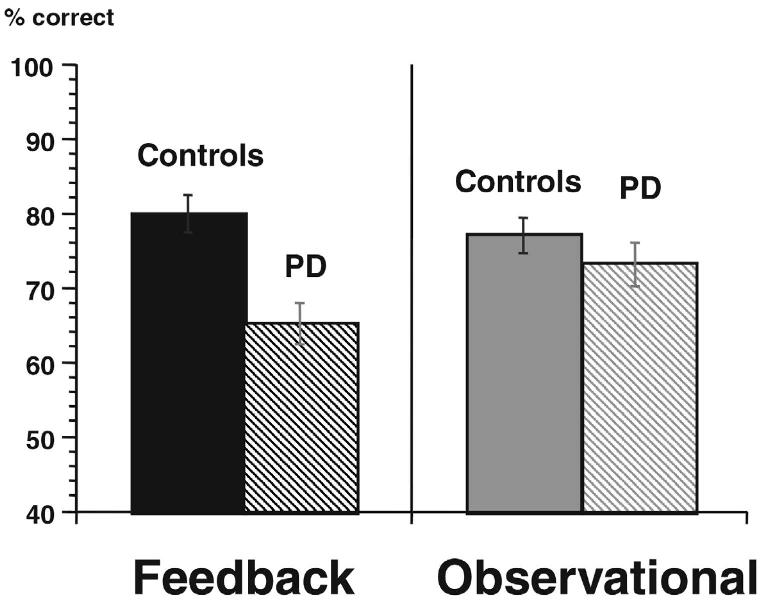


Figure 18–8. Percent correct for Parkinson's (PD) and controls on the observational task (right) and the test phase of the feedback-based task (left). Parkinson's patients are impaired at probabilistic category learning in feedback training but show no deficit when trained using observational training. Data from Shohamy, Myers, Onlaor, & Gluck (2004).

task, but were unimpaired when subsequently challenged to transfer what they had learned to a novel set of stimuli. The opposite pattern was found among individuals with hippocampal-region damage—spared initial learning, but impaired transfer. This suggests that PD patients do not have a general memory or cognitive deficit; rather, their deficit appears specific to the acquisition of cognitive skills through error-correcting feedback over many trials.

GENERAL DISCUSSION

In the late 1980s, Gordon Bower and I showed that there were common error correction principles for associative learning in both classical conditioning and probabilistic category learning, which allowed us to map the Rescorla–Wagner model of classical conditioning up to a larger-scale connectionist network model of human learning (Gluck & Bower, 1988a). In the intervening years, there has been significant progress in understanding the neural substrates of classical conditioning, implicating the cerebellum for error correction learning in aversive conditioning of motor reflexes, the basal ganglia for error correction learning for appetitive conditioning of reward-predicting stimuli, and the hippocampal region for supporting both forms of learning through modulation of the representations of stimuli that enter into these forms of learning. Integrating across both traditions, Catherine Myers, Daphna Shohamy, and I at Rutgers University-Newark (along with numerous collaborators at various other institutions) have shown in recent years that there are also common neural mechanisms for classical conditioning and human category learning, drawing on multiple methodologies in cognitive neuroscience, including neuropsychological studies of clinical populations and functional brain imaging. This provides a foundation for ongoing and future studies that seek further understanding of the cognitive neuroscience of human learning and memory, along with clinically relevant insights into neurological and psychiatric disorders that affect the basal ganglia (e.g., Parkinson's disease, Huntington's disease, and dystonia) and the hippocampal region (e.g., amnesia and Alzheimer's disease). The two strands of research begun at Stanford 20 years ago as independent avenues of inquiry into, first, human learning behavior and, second, the neural substrates of learning are now deeply intertwined into a single line of cognitive neuroscience research.

REFERENCES

Allen, M. T., Chelius, L., & Gluck, M. A. (2002) Selective entorhinal lesions and non-selective cortical-hippocampal region lesions, but not selective hippocampal lesions, disrupt learned irrelevance in rabbit eyeblink conditioning. *Cognitive Affective and Behavioral Neuroscience, 2,* 214–226.

Coutureau, E., Killcross, A., Good, M., Marshall, V., Ward-Robinson, J., & Honey, R. (2002). Acquired equivalence and distinctiveness of cues: II. Neural manipulations and their implications. *Journal of Experimental Psychology: Animal Behavior Processes, 28*(4), 388–396.

Daum, I., Schugens, M. M., Ackermann, H., Lutzenberger, W., Dichgans, J., & Birbaumer, N. (1993). Classical conditioning after cerebellar lesions in humans. *Behavioral Neuroscience, 107*(5), 748–56.

Donegan, N. H., Gluck, M. A., & Thompson, R. F. (1989). Integrating behavioral and biological models of classical conditioning. In R. D. Hawkins & G. H. Bower (Eds.), *Psychology of learning and motivation* (Vol. 23, pp. 109–156). New York: Academic Press.

Fera, F., Weickert, T. W., Goldberg, T. E., Tessitore, A., Hariri, A., Das, S., Lee, B., Zoltick, B., Meeter, M., Myers, C. E., Gluck, M. A., Weinberger, D. R., & Mattay, V. S. (2005). Neural mechanisms underlying probabilistic category learning in normal aging. *The Journal of Neuroscience, 24*(49), 11340–11348.

Gabrieli, J. D. E., McGlinchey-Berroth, R., Carrillo, M. C., Gluck, M. A., Cermak, L. S., & Disterhoft, J. F. (1995). Intact delay-eyeblink classical conditioning in amnesia. *Behavioral Neuroscience, 109*(5). 819–827.

Gluck, M. A., Allen, M. T., Myers, C. E., & Thompson, R. F. (2001). Cerebellar substrates for error-correction in motor-reflex conditioning. *Neurobiology of Learning and Memory, 76,* 314–341.

Gluck, M. A., & Bower, G. H. (1988a). Evaluating an adaptive network model of human learning. *Journal of Memory and Language, 27,* 166–195.

Gluck, M. A., & Bower, G. H. (1988b). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*(3), 227–247.

Gluck, M. A., & Bower, G. H. (1990). Component and pattern information in adaptive networks. *Journal of Experimental Psychology: General, 119*(1), 105–109.

Gluck, M. A., Bower, G. H., & Hee, M. (1989). A configural-cue network model of animal and human associative learning. In *Proceedings of the 11th Annual Conference of the Cognitive Science Society*, Ann Arbor, MI (pp. 323–332). Hillsdale, NJ: Lawrence Earlbaum Associates.

Gluck, M. A., & Myers, C. (1993). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus, 3*(4), 491–516

Gluck, M. A., & Myers, C. E. (2001). *Gateway to memory: An introduction to neural network models of the hippocampus and learning.* Cambridge, MA: MIT Press.

Gluck, M. A., Myers, C. E., & Thompson, R. F. (1994). A computational model of the cerebellum and motor-reflex learning. In S. Zournetzer, J. Davis, T. McKenna, & C. Lau (Eds.), *An introduction to neural and electronic networks* (2nd ed., pp. 91–80). San Diego: Academic Press.

Gluck, M. A., Oliver, L. M., & Myers, C. E. (1996). Late training amnesic deficits in probabilistic category learning: a neurocomputational analysis. *Learning and Memory, 3,* 326–340.

Gluck, M. A., Shohamy, D., & Myers, C. E. (2002). How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning and Memory, 9*, 408–418.

Gluck, M. A., & Thompson, R. F. (1987). Modeling the neural substrates of associative learning and memory: A computational approach, *Psychological Review, 94*(2), 176–191.

Hopkins, R. O., Myers, C. E, Shohamy, D., Grossman, S., & Gluck, M. A. (2004). Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia, 42,* 524–535.

Jarrard, L. E. (2002). Use of excitotoxins to lesion the hippocampus: Update. *Hippocampus, 12*(3), 405–414.

Kamin, L. (1969). Predictability, surprise, attention and conditioning. In B. Campbell & R. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.

Kim, J., Krupa, D., & Thompson, R. F. (1998). Inhibitory cerebello-olivary projections and blocking effect in classical conditioning. *Science, 279,* 570–573.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273,* 1399–1402.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning and Memory, 1,* 106–120.

Logan, C. G., & Grafton, S. T. (1995). Functional anatomy of human eyeblink conditioning determined with regional cerebral glucose metabolism and positron-emission tomography. *Proceedings of the National Academy of Sciences of the United States of America, 92*(16), 7500–7504.

Myers, C. E., & Gluck, M. A. (1994). Context, conditioning and hippocampal rerepresentation. *Behavioral Neuroscience, 108*(5), 835–847.

Myers, C., McGlinchey-Berroth, R., Warren, S., Monti, L., Brawn, C. M., & Gluck, M. A. (2000). Latent learning in medial temporal amnesia: Evidence for disrupted representational but preserved attentional processes. *Neuropsychology, 14*(1), 3–15.

Myers, C., Shohamy, D., Gluck, M., Grossman, S., Kluger, A., Ferris, S., Golomb, J., Schnirman, G., & Schwartz, R. (2003). Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *Journal of Cognitive Neuroscience, 15*(2), 185–193.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso-Moyano, J., Myers, C. E., & Gluck, M. A. (2001). Interactive memory systems in the brain. *Nature, 414,* 546–550.

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Rumelhart, D., McClelland, J., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (2 vols.). Cambridge, MA: MIT Press.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80,* 1–27.

Schultz, W., Dayan, P., & Montague, P. R. (1997) A neural substrate of prediction and reward. *Science, 275,* 1593–1599.

Sears, L., & Steinmetz, J. (1991). Dorsal accessory olive activity diminishes during acquisition of the rabbit classically conditioned eyelid response. *Brain Research, 545*(1–2), 114–122.

Shohamy, D., Myers, C. E., Geghman, K. D., Sage, J., & Gluck, M. A. (2006). L-Dopa impairs learning, but not generalization, in Parkinson's disease. *Neuropsychologia, 44*(5), 774–84.

Shohamy, D., Myers, C., Grossman, S., Sage, J., Gluck, M., & Poldrack, R. (2004b). Cortico-striatal contributions to feedback-based learning: Converging data from neuroimaging and neuropsychology. *Brain, 127*(4), 851–859.

Shohamy, D., Myers, C. E., Onlaor, S., & Gluck, M. A. (2004a). The role of the basal ganglia in category learning: How do patients with Parkinson's disease learn? *Behavioral Neuroscience, 118*(4), 676–686.

Squire, L. (1987). *Memory and brain*. New York: Oxford University Press.

Thompson, R. F. (1986). The neurobiology of learning and memory. *Science, 233,* 941–947.

Thompson, R. F., & Gluck, M. A. (1989). A biological neural-network analysis of learning and memory. In S. Zournetzer, J. Davis, & C. Lau (Eds.), *An introduction to neural and electronic networks* (pp. 91–107). New York: Academic Press.

Thompson, R. F., & Gluck, M. A., (1991). Brain substrates of basic associative learning and memory. In H. J. Weingartner & R. F. Lister (Eds.), *Cognitive neuroscience* (pp. 24–45). New York: Oxford University Press.

Trabasso, T., & Bower, G. (1964). Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology* (pp. 32–93). Stanford, CA: Stanford University Press.