CHAPTER 6

Procedural Learning II: Habits and Instrumental Learning

There is an old story about students in a psychology class who used what they had learned about conditioning to modify the behavior of their professor. In one version of this story, the professor typically stood behind a lectern at the right side of the room and gave his presentation. The students agreed in advance that whenever the professor moved even slightly to the left, they would show strong attention, nod in approval of the material, and jot down each item said. Conversely, whenever he stood behind the lectern, they would look around, yawn, and frown. Before long, the professor left his usual stance and began moving frequently to the left for short periods before returning to the lectern. As the presentation continued, he moved farther and farther to the left and no longer returned to the lectern. By the end of the class, he was standing in the left corner of the room, delivering his presentation without the notes he had left behind on the lectern.

THIS SORT OF behavior change could have occurred without the professor being aware that his behavior was being manipulated by the students. In this example, the professor's behavior was *instrumental* in producing its consequences (the "reward" of student attention). Therefore, this kind of learning is called **instrumental conditioning.**

Instrumental conditioning is pervasive. For example, child rearing practices carefully apportion rewards and punishments for good and bad behaviors. Do you remember being given a "gold star" for a good act or a "time out" for bad behavior? These reinforcements—that is, rewards and punishments—that occur immediately following particular actions are intended to alter the likelihood of those behaviors. The same ideas predominate methods of training puppies with a carefully designed mix of rewards and punishments that shape their behaviors to our liking. We use reinforcers in all kinds of real-life situations, shaping the behaviors of our significant others, our employees, and more.

Instrumental learning creates a particular form of memory called a **habit**, which involves voluntary behaviors that are brought about and influenced by reinforcers (rewards and punishments) presented predictably after the behaviors. In this chapter we will look closely at the brain system that supports the acquisition and expression of habits. We'll also discuss how habits and cognitive memory differ in both their properties and the brain systems that support them.

Fundamental Questions

1. What are the rules of instrumental conditioning?
2. Can we condition any behavior using rewards and punishments?
3. Where in the brain are habits stored?

Instrumental Learning Changes Reinforced Behavior to Reflect Memory

How does instrumental learning differ from classical conditioning? In classical conditioning, the unconditioned stimulus automatically or reflexively evokes a behavior. In contrast, instrumental behaviors are voluntary behaviors whose likelihood is influenced by **reinforcers**—that is, rewards or punishments such as the attention or inattention meted out by the professor's clever students. But distinguishing between reflexive and voluntary behavior can be tricky. For example, when animals are trained to run to the location of a reward, we have no way of knowing whether their behavior is voluntary or automatic.

The most concrete, dependable distinction between classical conditioning and instrumental learning is that in classical conditioning, the unconditioned stimulus always follows the conditioned stimulus *regardless* of whether the subject emits the conditioned response (the CR). By contrast, in instrumental conditioning, the reinforcer is presented *only* if the subject emits the desired response. So in Pavlov's protocol, the food is delivered to the animal at a particular time after the CS regardless of the animal's behavior. In instrumental conditioning, the animal is given the food only when it displays the behavior that is being conditioned to occur. Either way the animal seems to behave as if it figured out the connection. In classical conditioning, the CR *anticipates* the reinforcer: A tone that has been repeatedly followed by an air puff to the eye can elicit a blink before an air puff happens. In instrumental conditioning, the response *is emitted to obtain or avoid* the reinforcer. Thus in both situations, the subject associates an initial stimulus and the consequence it predicts. The main difference is whether the subject's behavior directly affects the occurrence of the reinforcer.

Given these basic differences in how classical and instrumental conditioning are accomplished, are their learning processes fundamentally the same or different? In the past there was considerable controversy over this question. Today distinctions between classical conditioning and instrumental learning have been overshadowed by similarities and differences in what is learned and what brain systems are involved in diverse learning situations. Instrumental conditioning is subject to many of the same classical conditioning phenomena described in the previous chapter—including conditioned excitation and inhibition, extinction and spontaneous recovery, stimulus generalization and discrimination, second-order conditioning, blocking, latent inhibition, and contextual conditioning.

Reinforcers Modify the Predictive Relationship between Stimulus and Response

The formal study of instrumental learning began with Edward Thorndike (1898), whose investigations of animal intelligence you may remember from Chapter 1. Thorndike placed cats in puzzle boxes. The box doors latched outside, and each box had a mechanism an animal could reach from inside, such as a chain to pull or a pole to push that would open the latch. Thorndike would offer food outside the box and observe animals' attempts to escape the box and obtain the food. Although Thorndike intended to examine the intelligence of animals in solving the puzzles, his detailed observations led him to conclude that the puzzle solutions were supported by reinforced habits rather than insight or intelligence. Initially animals displayed a variety of investigatory behaviors; eventually they would open the latch as if by accident rather than by insight or analysis. However, when an animal was retested in the same box, the behavior that opened the latch would occur earlier, so that the escape time diminished across trials.

Thorndike attributed the decrease in problem-solving time to incremental strengthening of a learned connection between the experimental situation and the behavioral response, with the reward reinforcing the successful behavior. He formalized this notion in his proposed **Law of Effect**, which says that if a particular response to a stimulus is followed by a "satisfying event," the satisfying event will strengthen the stimulus–response association, increasing the likelihood that the behavior will be repeated. On the other hand, if a particular response is followed by an "annoying event" (Thorndike's term), the annoying event will weaken the association, making the initial behavior less likely. Note that in Thorndike's conception, the reinforcing event itself is not part of the association. Instead the reinforcer only modifies the association between the stimulus and the response.

Reinforcement Alters Behavior in Appropriate Situations

Thorndike's experiments raised several issues that have been clarified in succeeding research. One of these issues is the nature of the relationship between behaviors and reinforcements. Thorndike proposed that contiguity between ongoing behavior and reinforcement is necessary and sufficient to produce learning. His Law of Effect predicted straightforwardly that any specific behavior occurring at the time of reinforcement will increase in likelihood. This central prediction of the Law of Effect has been validated many times in demonstrations of **superstitious learning**. Everyone has seen examples of superstitious behaviors—often little acts that people perform even when they are aware of no causal relationship between the action and its consequences. These behaviors typically are coincidentally associated with successful outcomes, are acquired rapidly, and can be persistent.

Some familiar examples of superstitious learning involve the behaviors of baseball players just before they bat. A particularly powerful example is the routine performed by Nomar Garciaparra, a well-known baseball player. Each time he comes to bat, he tightens his wrist bands repetitively several times, grinds his feet into place in a particular sequence, and then crosses his chest before taking position. Although some components of this behavior may actually influence his batting success, the ritualized nature of his routine is likely more superstition than crucial physical preparation.

Other strong evidence that simple contiguity of behavior and reward is sufficient for learning comes from B. F. Skinner's most famous experiment (Skinner, 1938). Skinner put a pigeon into a small enclosure fitted with a retractable food dispenser and arranged for food to be available at fixed time intervals. There was no predetermined relationship between any particular pigeon behavior and food availability. But despite the lack of an explicit contingency, whatever the pigeon happened to be doing when it received the food reinforcement subsequently increased in frequency. For example, one pigeon that had turned counterclockwise just before receiving food began frequently turning counterclockwise, presumably in hopes of receiving more food. One bird poked its beak into a corner of the test chamber, and another hopped around from one foot to the other. Skinner's explanation was that any behavior that was incidentally or accidentally occurring at the time of reward delivery would be reinforced.

Although Skinner's studies demonstrate the power of contiguity between behaviors and rewards, subsequent studies imply that not all behaviors are equally subject to the Law of Effect. In a key experiment, Staddon and Simmelhag (1971) repeated Skinner's study, more systematically and thoroughly examining the frequencies of a list of behaviors: turning, wing flapping, preening, movement along the chamber wall, orienting to the food dispenser, and pecking. All the pigeons performed certain behaviors, such as orienting to the food dispenser and pecking

along the neighboring chamber wall, more frequently near the end of the time interval, just before food was delivered. Thus not all behaviors were subject to the Law of Effect; only particular behaviors called **terminal responses** increased in likelihood when followed by a reinforcer.

Certain other behaviors, called **interim responses,** increased in occurrence in the middle of the interval between food deliveries. Staddon and Simmelhag suggested that these behaviors, including moving and turning along the food hopper wall, were related to searching when food delivery was unlikely. Timberlake and his colleagues have suggested that instrumental conditioning for food reinforcement should be interpreted in light of the feeding system that is activated in hungry animals (Timberlake & Lucas, 1989). Behavioral repertoires in this situation reflect preorganized, species-specific patterns of foraging and feeding. According to this model, when food is not available, animals use specific search patterns. After food delivery, they exhibit behavior focused on getting the food. From this perspective, the behavior patterns observed in these experiments can be viewed as learned alterations of the feeding system repertoire that correspond to reward delivery expectancies, rather than an increase in behavior that occurred contiguous with reward, as characterized by the Law of Effect.

A related set of findings concerns the extent to which particular types of responses can be instrumentally learned. In Chapter 5 we saw that some types of CS–US associations are readily conditioned and others are not. For example, a visual CS is easily conditioned to a shock US and food tastes are readily conditioned to illness; but it is difficult to condition visual cues to illness and tastes to shock (Garcia & Koelling, 1966). Similarly, some types of instrumental conditioning responses are more easily learned than others. Thorndike found that cats were not easily conditioned to execute some types of responses, such as grooming or yawning, to escape from a puzzle box. Based on these observations, he proposed the concept of **belongingness** (discussed in Chapter 5) to explain that certain responses "belong" with a particular reinforcer based on the animal's evolutionary history. Thus behavioral actions on objects a cat can manipulate, such as string pulling and latch pushing, belong with the movement of obstacles such as the puzzle box door, whereas yawning and grooming do not.

This principle was strikingly demonstrated in Breland and Breland's 1961 observations of the "misbehavior" of zoo animals they attempted to instrumentally condition. For example, a pig trained to drop a coin into a box tried to use modified versions of its natural rooting behavior to flip the coin into the air toward the box rather than adopting the seemingly more straightforward behavior of carrying the coin in its mouth and dropping it into the box. When they endeavored to train raccoons to perform a similar task, they were stymied by the raccoons' preoccupation with rubbing coins against one another and dipping them repeatedly into the box without dropping them. The investigators saw that the range of possible

instrumentally learned responses was limited by the subjects' natural repertoire of innate behaviors. In another systematic investigation of belongingness, Shettleworth (1975) identified six different natural hamster behaviors and attempted to increase the frequency of each behavior in different hamsters through instrumental conditioning. Food reinforcement was effective for behaviors hungry hamsters are likely to perform (such as digging in bedding and rearing and scrabbling at the walls) but not for nonfeeding behaviors (such as face washing, grooming, and scent marking).

Animals Develop Expectations about Behavior and Rewards

Thorndike proposed that instrumental learning involved an association between the stimuli present when the behavior was executed and the reinforced behavioral response, thus leading to the abbreviation *S–R learning*. In this conception, the reinforcer was not part of the association but simply strengthened the S–R bond. However, this original notion has been qualified by subsequent research. A series of influential studies showed that animals develop associations among all three relevant events—the stimulus, the response, *and* the reinforcer (Rescorla, 1988). One line of evidence is the observation that rats can learn different instrumental responses for different types of reinforcements. For example, a rat can be trained to press a lever for food and to pull a chain for water. So animals appear to associate different responses with specific rewards, and they seem to form expectations that particular responses will produce certain rewards.

Another line of evidence comes from studies in which the attractive value of a reinforcer changes after instrumental learning occurs. In this experimental process, called **reinforcer devaluation,** rats are initially trained to make two different responses for distinctive rewards, such as pressing one lever for food and another for sugar water. Subsequently, outside the learning situation, the animal is allowed to consume one of the reinforcers (perhaps the sugar water) and then is made sick by injection of lithium chloride. As described in Chapter 5, this results in classical conditioning of a taste aversion to sugar water. The key test comes next: examining the rate of lever pressing for food and sugar water following the specific devaluation of sugar water. If the rewards only *reinforce* the association between the appropriate lever (stimulus) and the lever press (response), then devaluing a particular reinforcer (sugar water) should have no effect. If, on the other hand, sugar water is part of the association formed between a particular lever and lever pressing, and the animal presses that lever in expectation of sugar water, we might expect a subsequent change in pressing on that lever. In fact, when sugar water is devalued, rats selectively avoid the water-associated lever and prefer the food-associated lever. This preference shows that the reward is indeed a critical part of the learned association.

Another type of experiment highlights the importance of animals' expectations by shifting reward magnitudes. Rats placed at one end of a simple linear alley receive food rewards when they reach the other end. Animals that receive greater amounts of food run slightly faster in the alley than those getting less food. However, if the amounts of food are shifted, the rats that ran quickly for large rewards will run more slowly than the rats that consistently received a small reward. Conversely, rats switched from small to large rewards will run more quickly than rats that always received large rewards. Thus the shift in expectations is more powerful than the actual amount of reward in controlling the animals' instrumental response.

A powerful and tragic demonstration of animal and human awareness of the contingency between behavior and reinforcement is **learned helplessness:** an acquired feeling of futility in which people believe they have no control over their situation. In a series of experiments by Martin Seligman and his colleagues (Seligman & Maier, 1967), animals were initially given painful shocks at unpredictable times. Other animals learned to perform a response that would allow them to avoid the shocks. Subsequently both groups of animals were put in a completely different situation where they could learn to avoid shocks by performing a cued response. The animals that had initially learned to avoid shocks readily acquired the new response. In contrast, the animals that had suffered inescapable shocks made no attempt to learn the new task; they had apparently learned there was nothing they could do to escape shocks.

Seligman extended the concept of learned helplessness to human studies. In one experiment he subjected college students to a series of unpredictable loud noises they could not control. Subsequently the students were asked to solve a series of anagrams, which students not exposed to the unpredictable noises could do well. The subjects exposed to unpredictable noise, however, had considerable difficulty in performing this task, even though solving anagram problems had little to do with the previous noise exposures. Seligman theorized that the earlier experience with uncontrollable unpleasant events produced a sense of helplessness and lack of control that carried over to performance on a subsequent unrelated cognitive task.

Learning & Memory in Action

What Is the Basis of Losing Streaks?

Based on his experimental analyses of animal and human learned helplessness, Martin Seligman (1975) suggested that depression (a lack of affect and feeling) may be attributed to a person's learning that he or she has no control over events. Thus a sudden life-changing illness, such as

cancer, or a disaster like a flood may be considered a negative reinforcer that is unpredictable and inconsistent with behaviors preceding these events. Under these circumstances, some people can develop a deep sense of loss of control that extends broadly to how they approach not only their specific challenge but also many other life problems. Seligman described the case of a woman whose children had gone to college and whose husband traveled extensively. She felt that these important events were outside her control and developed profound depression.

The idea of learned helplessness has been extended in several directions to elucidate various phenomena. For example, some people have suggested that losing streaks by sports teams may be the result of learning that aversive events cannot be controlled. One study found that indeed teams that lost badly one week were more likely than predicted by their overall performance to lose the next week. Conversely, is it possible that winning provides a false but effective sense of control? How else can anyone explain how the 2004 Red Sox managed to beat the Yankees following a mounting series of successful events during the divisional playoffs?

A practical application of learned helplessness may explain poor classroom performance by some children. Early unexplained failures may create a sense that study and homework behavior does not control outcome. If students view their grades as unpredictable, they may assume they will do poorly regardless of their efforts. Some programs attempt remedial tutoring based on clear assignments, strategies to achieve success, and rewards based on following the specific assignments to reverse such feelings of helplessness and turn around otherwise unsuccessful classroom performance.

Interim Summary

Thorndike's Law of Effect states that reinforcers alter the likelihood of behaviors that precede them. This simple rule explains the development of superstitious behaviors that we acquire because of their history of association with reinforcing outcomes. However, the Law of Effect depends on the types of behaviors, called terminal responses, that tend to occur just before a reward is consumed, as opposed to interim responses that occur preceding reward consumption. And not all behaviors are equally easily conditioned by reinforcers. Rather, behaviors that "belong" with particular reinforcers (like foraging behaviors that precede food) are more easily conditioned by an instrumental contingency. Consistent with this observation, animals develop expectations about rewards that follow their behaviors and act accordingly. For example, they modify their instrumentally conditioned responses appropriately

when reward values are altered. Also, behavior can be affected by a learned absence of reinforcer predictability. In particular, animals and humans become "helpless" when conditioned to believe that negative reinforcements are unpredictable and unrelated to their behavior.

Animals Learn about the Environment and Expect Reinforcers

A variety of protocols are used to study instrumental conditioning. Among the most common are mazes and Skinner's operant box. Each of these learning situations has provided deep insights into the processes and neural bases underlying instrumental learning.

Rats in Mazes Learn Rules, Expectancies, and Spatial Layout

In 1901 Willard Small, inspired by the famous garden maze at Hampton Court in London (Figure 6.1), introduced the maze to studies of animal learning. He began what would become an industry of systematic and quantitative studies aimed at identifying the minute details of how rats acquire specific turning patterns in mazes. Some mazes used to study rats' navigational strategies are simple. Perhaps the most common example is the T-maze (Figure 6.2, left), which simplifies maze

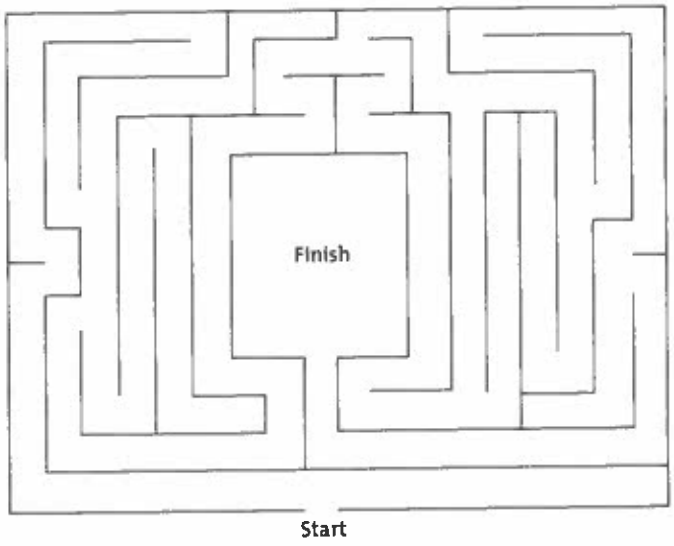

FIGURE 6.1 The maze in Hampton Court.

navigation to a single turn. Rats begin at one end of an alley and at the other end reach a choice point between right and left goal alleys. In one common test condition, the rat is required to make a particular left or right turn consistently. This condition is called "win–stay" because the rat "wins" a reward by turning in one direction and reaching a particular goal arm, and it must "stay" with that response in subsequent trials to obtain additional rewards. A variation is the "win–shift" version of the task: After "winning" a reward in one goal arm, the rat must then "shift" to the opposite goal arm to receive the next reward. In *alternation learning* the rat continues to "win–shift" over many successive trials, thus alternating between turns to the left and right goal alleys.

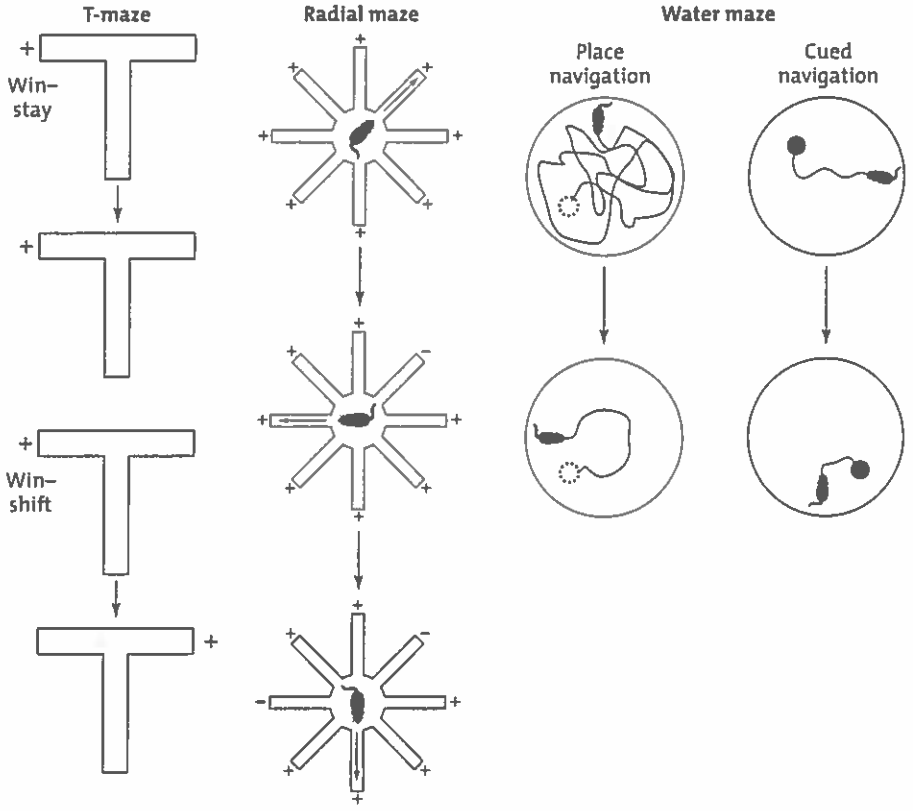

FIGURE 6.2 The radial maze, T-maze, and water maze.
In the T-maze win–stay task, the reward site (+) is consistent across trials; but in the win–shift task, its location is switched between trials. In the radial maze, rewards are found only once at the end of each maze arm. In the water maze place navigation task, the escape platform cannot be seen and is always placed in the same location. In the cued navigation version of this task, the platform can be seen and is moved from trial to trial.

An animal's ability to learn expectancies about rewards in maze learning is seen in comparisons between win–stay and win–shift performance: Rats more readily learn the win–shift task than the win–stay task. This result may seem counterintuitive because the animal most readily learns to *not* repeat the most recently rewarded response. Certainly this result is not predicted by the Law of Effect. It appears that after the rat has eaten the food in one goal arm, it does not expect to find food there on the next trial and instead searches the other goal alley.

The most amazing performance of win–shift behavior is observed in the radial maze, composed of a central platform with multiple goal arms radiating like spokes of a wheel in all directions (Figure 6.2, center). In the win–shift version of this task, each goal arm end is baited with a food pellet, and the rat is allowed to forage. Rats readily learn to visit each arm just once in a trial, collecting all the rewards efficiently by avoiding arms already visited. The rats could have learned to visit the arms in a succession of clockwise or counterclockwise choices. However, the procedure delays each choice by a few seconds, which is sufficient to prevent that strategy. Instead the animals seek food in new maze arms in a seemingly random order without returning to previously searched areas. Rats performed surprisingly well on this task even when there were 17 goal arms, requiring the rat to remember as many as 16 arms visited in each trial (Olton et al., 1977).

Another popular maze task is the water maze (Morris, 1984)—an open swimming pool filled with warm water that is made opaque with the addition of milk powder or water-based paint (Figure 6.2, right). There is no way out of the pool, but a rat can climb on a platform in the water to rest. Across a series of training trials, the rat begins from any of several starting locations and must find the platform to climb on it. In one version of the task, called *cued learning*, the platform is elevated just above the water surface and painted with a dark color so that it is easily visible. The painted, raised platform is usually moved between trials so the rat must learn to approach the visible cue rather than swim to the former location of the platform. Here a simple S–R association between a specific visual cue and the approach response toward that cue supports learning.

In another version of the task, called *place learning*, the platform is set just below the surface of the water with no cue indicating its location. Rather, rats must learn to use environmental cues to remember the platform location. Here learning requires more than a simple S–R association. Because no specific cue indicates the platform's location, the rats cannot simply learn to approach a specific stimulus. And because the animals must swim from different starting points, they cannot learn a specific set of turning responses to reach the platform.

The other common paradigm for studying instrumental learning is the **free operant** technique developed by Skinner. The term *operant* refers to the requirement that the animal must operate on the environment to receive a reward. Thus operant learning is generally viewed as synonymous with instrumental learning.

Skinner developed the free operant protocol to let the subject control when it executed the response without experimenter intervention. In a typical free operant task, a rat presses a lever or a pigeon pecks a key, and at any time that behavior increases the likelihood of reward. The measure of learning is a change in response rate associated with reward contingency. Mazes also use operant conditioning in the sense that animals' operating responses to the maze environment get specific reinforcements. However, maze experiments involve discrete trials in which the experimenter imposes conditions for when the instrumental behavior is to be executed. Also, the typical measure of learning in a maze is accuracy of choice and the latency to make the choice once the trial has begun, rather than the response rate. The following sections will describe how these paradigms are used to study instrumental learning.

Operant Conditioning Shapes Behavior to Match Rewards

Special techniques have been developed to study operant conditioning. The apparatus used for operant conditioning is typically a small chamber (often called a **Skinner box** after its creator) outfitted differently depending on experimental subject species (Figure 6.3). Rats press a lever to get food pellets delivered into a cup from a gadget that looks and works like a gumball machine. Pigeons have a round panel in the wall, called a key, that they can peck to obtain grain.

In operant conditioning animals typically learn uncommon behaviors, such as a rat pressing a lever or a pigeon pecking a key. The timing of reward consumption is often a bit delayed from the desired operant behavior; even if food is delivered immediately after a lever press or key peck, it is presented a few inches away from the lever or key, and the animal may not even notice the food itself while making the operant response. Because of these features of operant conditioning, a **shaping** procedure gradually modifies the animal's behaviors toward the desired responses. In shaping, the experimenter begins by delivering a reward the first time the animal moves close to the feeder. When the animal first hears the feeding machine deliver the food, it may freeze at the noise. However, eventually the animal finds the food and becomes interested in the feeder. After a few food deliveries, the animal habituates to the noise and (by classical conditioning) associates the noise with the appearance of food in the feeder. At this point the feeder noise has become a conditioned stimulus and is referred to as a **secondary reinforcer.** It takes on a key property of food (the primary reinforcer): It entices the animal to approach the feeder.

Now the experimenter shapes the animal's behavior so that it will perform the desired response. This is accomplished by rewarding successive approximations of the desired behavior. At first the experimenter delivers the food when the animal merely moves toward the lever or key. When the animal increases its approach

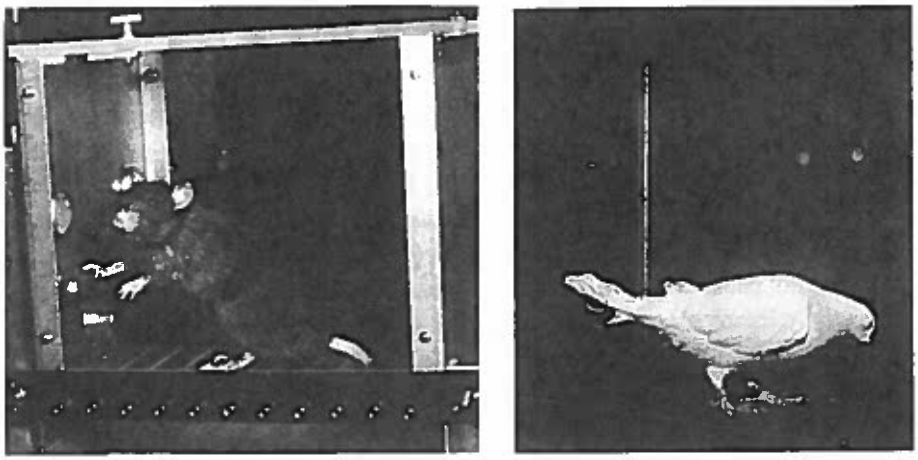

FIGURE 6.3 The Skinner box.
Left: Rat in lever-pressing test. The lever is just to the left of the rat. Lights in the box signal the availability of food when the lever is pressed. Right: Pigeon in key-pecking test. The pigeon will peck at one of the three keys to generate a food reward that will be available in the hopper located lower on the wall.

behavior, the experimenter requires the animal to perform the next component of the conditioned behavior: The rat must contact the lever or the pigeon must peck near the key before being rewarded. The experimenter continuously requires a response closer and closer to the full lever press or key peck that automatically activates the feeder. This procedure resembles the strategy used by the students mentioned at the beginning of the chapter, who shaped their professor's lecture behavior so he would stand in one corner of the room.

An interesting variation on this procedure is **autoshaping**. In this protocol, untrained pigeons are placed into the conditioning chamber. At a regular interval, the response key is illuminated for several seconds, then food is delivered. As in the conditioning of superstitious behavior, there is no required relationship between the desired key peck response and the food delivery. Nevertheless, pigeons reliably begin to peck the key when it is illuminated before the automatic food delivery; at that point the feeder activation is made to require a key peck. The usual interpretation of this phenomenon is that early in conditioning, the pigeon looks toward the suddenly illuminated key before reinforcement. This association may reinforce the looking behavior. Later, when food is not delivered immediately upon looking, the pigeon begins to approach the key, and it does this ever closer to the time of food delivery. According to this view, in later trials, as the latency between looking at and approaching the illuminated key decreases,

the animals come even closer to the key and begin to peck at it; that behavior becomes associated with the reward.

However, careful observations of the pigeons show that this interpretation is not correct. Pigeons do not successively look, approach, and contact the key. A study that closely examined the pigeons' behavior at the time of pecking hinted at the actual mechanism of autoshaping (Jenkins & Moore, 1973). Here some pigeons were trained to peck for food and others for water. Detailed observation showed that the pecks differed under these conditions. When the reinforcer was food, the pigeons executed brief, forceful pecks with their beaks open, just like those performed when feeding. When the reinforcer was water, the pigeons instead pecked slowly with their beaks closed and even sometimes made swallowing movements, just like drinking behavior. These observations have led to the interpretation of autoshaping as an example of classical conditioning. According to this view, movement toward the key replicates the unconditioned behavior that was executed when the reward was delivered; the illuminated key becomes a CS as it comes to substitute for the food (US), taking on a similar ability to elicit feeding behaviors. Autoshaping is therefore an example of how instrumental and classical conditioning share many features.

The field of **behavioral modification** applies principles of shaping and instrumental or operant conditioning to a broad variety of situations in education, business, and psychotherapy. These methods can diminish problematic behaviors and increase appropriate behaviors of children in the classroom; teach children with learning disabilities to speak and write more fluently; reduce workplace accidents; and help people lose weight, reduce alcohol consumption, and improve their lives in many other ways.

One example is the use of **token economies** in a variety of situations including classrooms, mental institutions, and prisons. In a token economy, individuals earn tokens for performing or not performing specific behaviors. The tokens can be exchanged for primary reinforcers (such as candy or other desirable goods). In a mental institution, tokens may be given for personal hygiene, appropriate social interactions, participation in group discussions, and productive work activities. These therapies succeed to the extent that specific behaviors can be defined and consistently reinforced with tokens. In some situations a token economy has succeeded as well as other traditional therapies and has reduced the need for medication. Token economies are also often used in classrooms to encourage good behavior and stronger academic performance.

Reinforcement Schedules Determine Behavioral Responses

Real-world reinforcements do not immediately or reliably follow behavior. For example, we often work hard in studying for a test but receive the grade substantially after taking it; and sometimes we do not receive the good grade we expect. Gamblers pay to pull a slot machine lever many times to obtain infrequent rewards. Baseball batters work hard to get a hit about a third of the time. Why

do we continue to produce learned behaviors when the reinforcement contiguity is delayed or infrequent? What controls when we continue to generate the behavior and when we quit (extinguish)?

A simple, striking illustration of the impact of irregular response–reinforcement contiguity is the phenomenon called the **partial reinforcement effect.** Animals' behavior is first shaped with rewards reliably following each response. But once the behavior is engaged, we can easily switch to reinforcing only every other response; and we can gradually decrease the reinforcement to only once every 10 responses (or even fewer) and still maintain robust responses. Indeed, under these conditions, when the reinforcement is discontinued, the number of trials required before the animal ceases responding (extinguishes) is considerably *greater* than the number of trials to extinction if every response was rewarded before extinction.

In daily life, partial reinforcement sometimes works against our best intentions. For example, a parent who punishes or tries to extinguish a child's whining for a treat should, in principle, be effective. However, if the parent occasionally gives in and rewards the child, this produces a variable ratio schedule (explained in the following paragraphs) that extends rather than reduces the behavior.

The partial reinforcement effect may seem paradoxical because the learned behavior seems to be stronger for the less rewarded condition—an apparent violation of the Law of Effect. This view comes from the observation that the contingency of partial reinforcement is similar to that of extinction: In both conditions, the subject emits many nonreinforced responses. However, in partial reinforcement the subject may associate many nonreinforced responses with eventual reinforcement. Thus, following partial reinforcement, the subject may continue to respond in extinction because it has learned to expect that one of its responses will eventually pay off.

Much research has focused on the relationship between operant behaviors and patterns of partial reinforcement in an effort to understand how behavior is modulated by unreliable rewards. Four reinforcement schedules have received the most study. In a **fixed ratio schedule (FR),** reinforcement is given after a particular number of responses. For example, in an FR-4 schedule, reinforcement is delivered after each fourth response. In a **variable ratio schedule (VR),** the number of responses required before the reward varies randomly but has a constant average value. In a **fixed interval schedule (FI),** reinforcement is delivered following the first response after a particular amount of time since the last reinforcement. For example, in an FI-15 schedule, a timer begins after each reinforcement delivery; when it has measured 15 seconds, another reinforcement is given after the next response. Finally, in a **variable interval schedule (VI),** the duration of the minimum interval varies randomly but has a constant average period. For example, in a VI-15 schedule, the time before the next response generates a reward might vary between 5 and 25 seconds, with an average of 15 seconds.

Each of these schedules results in a distinct behavior pattern that is typically recorded in terms of the running sum or **cumulative response** over time (Figure 6.4). Gener-

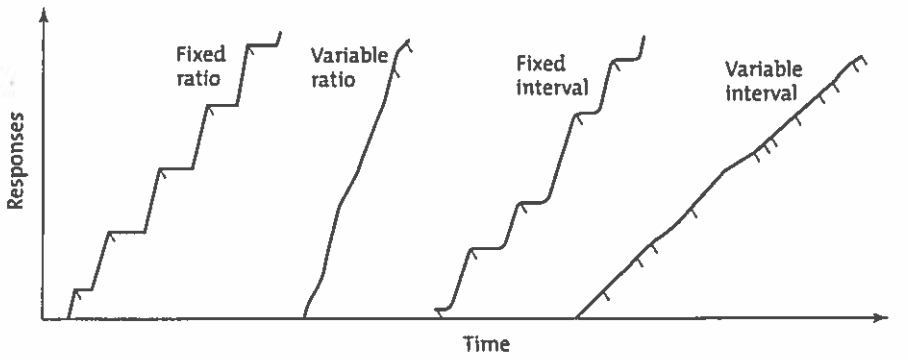

FIGURE 6.4 Operant response patterns for different reinforcement schedules. The *x* axis represents time, and the *y* axis represents the cumulative number of responses. The small marks show when reinforcements are provided.

ally response rates are higher in the ratio schedules than in the interval schedules, resulting in a more rapidly rising cumulative response. In fixed ratio schedules the subject usually pauses briefly after each reinforced response; but the cumulative response is more or less linear for both fixed and variable ratio schedules. In a fixed interval schedule, the cumulative response has a scalloped appearance, with an initially slow response rate and then increasing response as the interval times out. This response pattern suggests that the subject anticipates the reward with an increasing rate of behavior. In a variable interval schedule, the time of reward cannot be judged; here the cumulative response is linear, indicating a constant response rate.

Considerable research has examined the response pattern when an animal can choose between two behaviors (usually two levers to press or two keys to peck) that are associated with different reward schedules. For example, a rat might be allowed to respond to a left lever on a VI-15 schedule and a right lever on a VI-30 schedule; that is, rewards can be obtained at a minimum of every 15 seconds on the left lever and 30 seconds on the right lever. After substantial experience, the rat emits a consistent pattern of responses at the two levers to maximize the rewards. This pattern of behavior is described by the **matching law.** The response rates are described by this formula:

$$B1/(B1 + B2) = R1(R1 + R2)$$

B1 = the rate of behavior 1 (left lever presses)
B2 = the rate of behavior 2 (right lever presses)
R1 = the maximal rate of rewards associated with behavior 1
R2 = the maximal rate of rewards associated with behavior 2

In the example, R1 = 4 per minute and R2 = 2 per minute, so $B1/(B1 + B2) = 2/3$. In other words, the rat will distribute its lever presses between the two levers

at a 2:3 ratio, with a greater rate at the left lever. The same formula accurately characterizes choice behavior for variations in the magnitude of reward associated with two behaviors.

Note that this matching law will maximize the reward rate only in a combination of variable interval schedules where an exact interval length cannot be anticipated. In a combination of fixed interval schedules, responding to one lever will always provide a greater amount of reward than responding to the other; so animals will eventually respond only to the lever with the higher rate of reward.

Interestingly, in similar situations humans are sometimes characterized as irrational when they do not consistently select the higher-probability choice but instead pursue what is called **probability matching**. In probability matching subjects select each of several choices in proportion to their rate of success. Thus, in typical experiments, subjects choose between two buttons and are rewarded for predicting which button will illuminate next. Most humans will match their choices to the probabilities of the two buttons becoming illuminated. For example, if the buttons pay off in a ratio of 2:1, human subjects tend to select the buttons in the same ratio even though they would be correct more often if they would consistently select the higher-probability button. In contrast, animals do not probability match; rather, they consistently select the higher-paying button. Sometimes animals really are smarter than people!

Chained Responses Compose Organized Behavioral Routines

We have seen so far that animals can exhibit seemingly intelligent behavior in responding to different reinforcement schedules and in making choices. Here we will extend the scope of instrumental behavior to sequences of responses and stimuli. A straightforward extension of simple instrumental responding is the phenomenon of **response chaining**—which is how animals execute a sequence of operant behaviors, only the last of which results in reward presentation. Some prominent examples of response chaining are the acts animals perform in circuses or similar public shows. Trainers can teach an animal to climb a ladder, then pull a chain, run through a tunnel, push a ball to a cup, slide down a chute, and finally run to the trainer for a reward. How are such complex behavior sequences maintained? The general explanation is that each behavior in the sequence serves as a secondary reinforcer to the next behavior, until the animal finally receives the primary reinforcer (food) from the trainer at the end of the sequence. In other words, completion of the first behavior is reinforced by the stimulus on which the animal acts in the next step of the sequence; the second act is reinforced by the stimulus for the third act; and so on.

But how does the trainer establish the indirect relationship between the initial act in a long sequence and the receipt of primary reinforcement at the end of the sequence? What we never see is that animals most effectively acquire response

chains using **backward chaining.** In this strategy, animals are first trained to perform the *final act* in the sequence to obtain the primary reinforcer. Then they are trained to perform the next to last act to get the stimulus for the final act, which results in reinforcement. Next they are trained to perform the third to last act to gain presentation of the next to last stimulus, and so on. Eventually the animal will perform the entire chain even if all the stimuli (the ladder, chute, and so on) are present at once. Response chains are fragile: If one of the stimuli is missing, the chain may be broken so the animal cannot skip a missing part of the sequence.

A related problem is the learning of **serial patterns.** Can animals discover sequential patterns in rewards and respond appropriately? Hulse and Campbell (1975) trained rats to run down a simple alley to receive varying numbers of food pellets in a reliable order: 14 on the first trial, 7 on the second, 3 on the third, and none on the fourth. The animals adjusted their running speed as if they could predict the anticipated reward, running most quickly on the first trial and slowing progressively to the last. A subsequent study showed that the rats were not simply running more slowly as the day went on. They could also learn to match running speed to reward expectation even if the sequence involved both increasing and decreasing reward magnitude changes (such as 14, 1, 3, 7, 0). One explanation is that rats might use each item in the sequence as the cue for the next, similar to the associations that underlie response chaining (Capaldi et al., 1980). However, Roitblat and colleagues (1983) found that inserting new nonrewarded responses in the middle of a sequence did not disrupt sequential pattern responding for the rest of the sequence. Thus the animals were not relying on each individual association between adjacent items, but must instead have developed knowledge about the overall positions of items in the sequence.

Interim Summary

The performance of rats in mazes has been used extensively to examine how rats learn and remember particular behavioral responses and locations in space. Variants of the popular T-maze protocol include the win–stay rule (rats are consistently rewarded for one type of turn), the win–shift rule (they are consistently rewarded for turning in the direction opposite that of the last response), or alternation (they systematically switch between left and right turn responses). The radial maze has several arms radiating from a central platform; training for this maze most commonly uses the win–shift rule. The water maze is a large pool filled with opaque water with an escape platform submerged at a particular location defined by cues outside the maze.

Other popular protocols for studying instrumental learning involve operant conditioning tasks in which animals are gradually trained to perform an otherwise low-frequency behavior, such as pressing a lever. Behaviors most related to

the one occurring just before reward consumption can be autoshaped by drawing attention to the location of the desired response before the reward is delivered.

Partial reward schedules make behaviors resistant to extinction. Typical partial reinforcement schedules can be fixed or variable and involve either the ratio of responses to rewards or the interval between reinforced behaviors. Also, when put into conflict situations where rewards are distributed between two behaviors, animals tend to select the more rewarding behavior alone, whereas humans tend to match their responses to the probability of rewards for each behavior. Instrumental learning can also support the chaining of responses into a long series of behaviors that forms a routine.

Humans' Habits and Skills Combine Cognitive Memory and Instrumental Learning of Motor Programs

The previous discussion suggests that animals can acquire sequential knowledge and behavior that mimic humans' complex skill learning. But how do people learn complex skills and sequences? Much effort has gone into the study of how humans learn to complete sequences and solve other complex skill-learning problems. Some forms of sequence learning that have received attention are straightforward examples of motor coordination such as walking, swimming, writing, typing, playing the piano, and driving a car. Walking and swimming are repetitive and the others not; but all these skills require precise timing and orderly actions. Without that precision, we would fall down when trying to walk and never play a pleasing tune on the piano. How do we become skilled at these procedures, and how do we seem to execute them so effortlessly?

An early view was that skilled motor behaviors are mediated by response chaining mechanisms. For walking, it seems reasonable that the sight of the left leg moving forward and stepping down might act as the stimulus for beginning to raise the right leg and move it forward. We could break these alternate leg movements into simpler elements of muscle contractions that follow in sequence. Those microscopic movements would eventually have to be accomplished unconsciously using sensory feedback from angles of the leg parts and the pressure of stepping down. Also, you can imagine how simple feedback reinforcement might shape and optimize a broad variety of coordinated behaviors. Improvements in timing and positioning of each act in the sequence would result from feedback of our smoothness and speed in executing the sequence.

However, as Karl Lashley (1951) realized, there are several reasons why response chaining does not adequately explain skill learning. First, reaction times are much too slow to have feedback from one movement initiate the next. The

time it takes a person to react to the sensation of his or her own movements is longer than 100 milliseconds. Yet a pianist can execute 16 finger movements in a second! Second, some people who have lost sensory feedback from their hand movements due to nerve damage can still execute skilled hand movements, even while blindfolded to prevent visual feedback. Third, many of the most common errors made in skilled performance involve sequencing errors, such as when we transpose sequential letters when typing (*raed* instead of *read*). This would seem impossible if each typed letter was the essential stimulus for the next. Fourth, it has been observed that the time required to begin a sequence depends on the length of the sequence to be executed.

These findings led to the notion of a **motor program**—a preprogrammed sequence that is initiated complete with sequencing and timing of its elements. For typing, a motor program would send all the commands for finger movements, initiating them in the appropriate order to produce the correct sequence of typed letters. An advantage of motor programs is that they can complete sequences at any speed. Modern views of motor skill learning contend that the motor program contains a sketch of the general coordination of movements that can be modified as needed to alter the tempo, magnitude, or intensity of the movements or even the specific body parts that are executing the movements (Schmidt, 1988). Examples of this occur when we make fine adjustments to improve the accuracy of our tennis serve, when we change the size of our handwriting, or when we alter the loudness or tempo in playing music on the piano.

Skills Are Learned in Three Stages

More complex forms of human skill learning are accomplished in a series of stages. Anderson (1982) proposed three characteristic stages of the development of a complex skill. The kinds of skills that require these stages are complex, such as learning to type or to drive a standard transmission car. The first **cognitive stage** involves the learner remembering a list of instructions for the sequence to be followed. For example, in learning to type you might think about the word *read*, slowly look at each key on the keyboard, and type the letters one at a time. In driving you might memorize the steps of engaging the clutch, then shifting, then pressing the gas petal, then releasing the clutch. This stage involves cognitive memory, and it is slowly and often awkwardly executed.

The second **associative stage** is also slow. This stage involves substituting for the verbal memory a more direct representation of the sequence of movements that executes the skill. At this stage in learning to type or drive, you become free of having to remember the next step in the sequence. Each step begins to smoothly follow the last. You no longer have to rehearse; you begin to perform the skill without conscious memory. The skill parts have become linked in a procedure

that still requires close attention and monitoring—but not anticipatory, conscious, step-by-step commands. You must think about the word to be typed or when shifting must occur, but directions for each sequential step are executed without recalling each step.

Finally, in the **autonomous stage**, the procedure is free of conscious control and direction. You can type by just seeing words, and your hands effortlessly press the correct keys. Similarly, all aspects of driving a car are executed smoothly and virtually outside conscious control. An expert typist can keyboard a page of text in seconds but have almost no idea of its content. Experienced drivers can carry a conversation or think about what they will do at their destination with almost no monitoring of their driving actions. When a skill is fully autonomous, explicit remembering can become difficult. Some expert typists find it hard to describe where some keys are on a keyboard. Seasoned drivers can have trouble telling someone how to change gears in a car—it's easier just to show them. An autonomous skill requires less attention but is less interruptible. It is difficult to type part of a common word or to stop at a particular move in shifting gears.

Interim Summary

We learn a variety of skills and routines in everyday life, including walking, writing, typing, and playing the piano. Some repetitive movements, like walking and swimming, may be learned by response chains in which each sequential action is reinforced by feedback of its success. Other sequential behaviors, such as writing, typing, and piano playing, involve motor programs that occur too quickly for feedback to have substantial influence; these programs can be fine-tuned in amplitude and tempo. Complex skills are learned in a series of stages. In the cognitive stage a learner remembers a list of steps; the associative stage eliminates explicit remembering but still requires conscious attention and monitoring; and the autonomous stage allows the skill to be expressed without conscious control.

Striatal Cortical Pathways Support Instrumental Learning and Skill Acquisition

What brain circuits support our capacity for learning habits and skills? Considerable evidence indicates that skill learning is accomplished by a brain system centered in the striatum, a brain area that was introduced in Chapter 1 (Figure 6.5). The striatum receives input from the entire cerebral cortex, and these connections are capable of neural plasticity. The striatal connections are organized to sort and associate somatosensory representations with the appropriate motor representa-

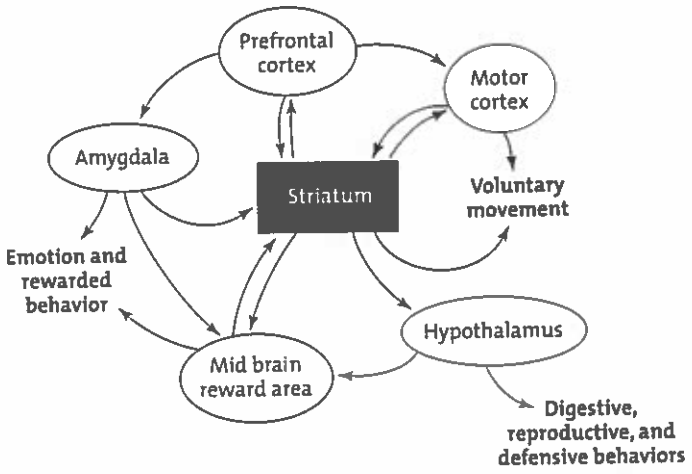

FIGURE 6.5 The striatum as a central structure in instrumental behavior.

tions—such as associating sensory representations of the hand with motor representations that control it (Graybiel et al., 1994).

The striatum is a part of a cortical loop that communicates with nuclei in the thalamus, targeting both the premotor and motor cortex and the prefrontal association cortex. Notably, this circuit projects only minimally to the brain stem motor nuclei and not at all to the spinal cord. In addition, the striatum interacts with many different brain areas involved in a variety of processes such as motivated (food seeking, defensive, and reproductive) behaviors; emotion; and processing rewards. This circuitry shows that the striatum does not directly control motor output but instead contributes to higher motor functions—including, many believe, the planning and execution of goal-oriented behavior (Graybiel, 1995; Schultz et al., 1995; Mink, 1996).

The Striatum Is Critical in Animal Instrumental Learning

Much research has suggested that the striatum plays a pivotal role in S–R learning (Packard & Knowlton, 2002). Evidence for this hypothesis comes from studies of the effects of striatum damage or inactivation in animals. Early studies found that striatum damage impaired learning of a variety of instrumentally conditioned tasks, including conditioned avoidance and discrimination learning tasks based on training with reinforcements.

How do we know that striatum damage specifically affects instrumental learning and not other types of learning or perception, motivation, or motor coordination? This problem has been addressed using **double dissociation,** in which the

experimenter typically compares the effects of damage in two brain areas on tasks that demand the same general perception, motivation, and motor coordination but distinct forms of learning and memory. Damage to brain area A affects performance on task 1, but performance on task 2 is normal; damage to brain area B affects performance on task 2, but performance on task 1 is normal. This pattern of results shows that both variants of the task are sensitive to brain damage; that neither locus of damage affects the common perceptual, motivational, or coordination requirements of the tasks; and that each locus of brain damage affects how these capacities create a distinct type of learning.

One striking double dissociation experiment involved training rats with a radial maze (eight arms radiating from a central platform), comparing the win–stay and win–shift rules you read about earlier in this chapter (Packard et al., 1989; Figure 6.6). Both versions of the task used the same maze, the same food-motivated performance, and the same kind of approach responses. The tasks differed only in the behavioral choice necessary to earn a reward, allowing researchers to examine the role of specific brain areas in different kinds of learning. In the win–stay task, four of the eight maze arms were illuminated each day; the rat could consistently approach any lit arm to receive a reward. Rats with striatum damage were impaired in learning this version of the task, but rats with hippocampus damage actually performed better than normal animals. In the win–shift task, all maze arms contained a reward each day, and the rat had to visit each arm just once and then shift to a new arm to maximize its rewards. (When the rat removed food from a particular maze arm, the reward was not replaced.) Rats with hippocampus damage were impaired in this task, but rats with striatum damage performed normally. The win–stay version of the task using illuminated maze arms can be viewed as a prototypical S–R task in that animals must respond consistently to a particular cue (arm illumination). This type of memory depends on the striatum (and not the hippocampus), as the rats' behavior showed. In contrast, the win–shift version of the task requires remembering which maze arms still have rewards and does not support performance based on a consistent S–R association. Instead it requires animals to remember which arms they have already visited; this form of memory apparently depends on the hippocampus and not on the striatum.

Perhaps you recall from Chapter 1 an elegant double dissociation of striatum and hippocampus memory functions. In that study, either the striatum or the hippocampus was temporarily inactivated to determine the roles of these brain areas in different learning strategies for a simple T-maze task. Rats were trained to make a consistent turn to enter a particular goal arm, then tested after one and two weeks of training with the maze rotated 180 degrees. After one week of training, rats tended to go where they had previously received rewards even though they had to turn in the direction opposite that used during training. This *place learning* depended on the hippocampus, not the striatum. In contrast, after two weeks of training, the rats tended to make the same turn they had used during training even

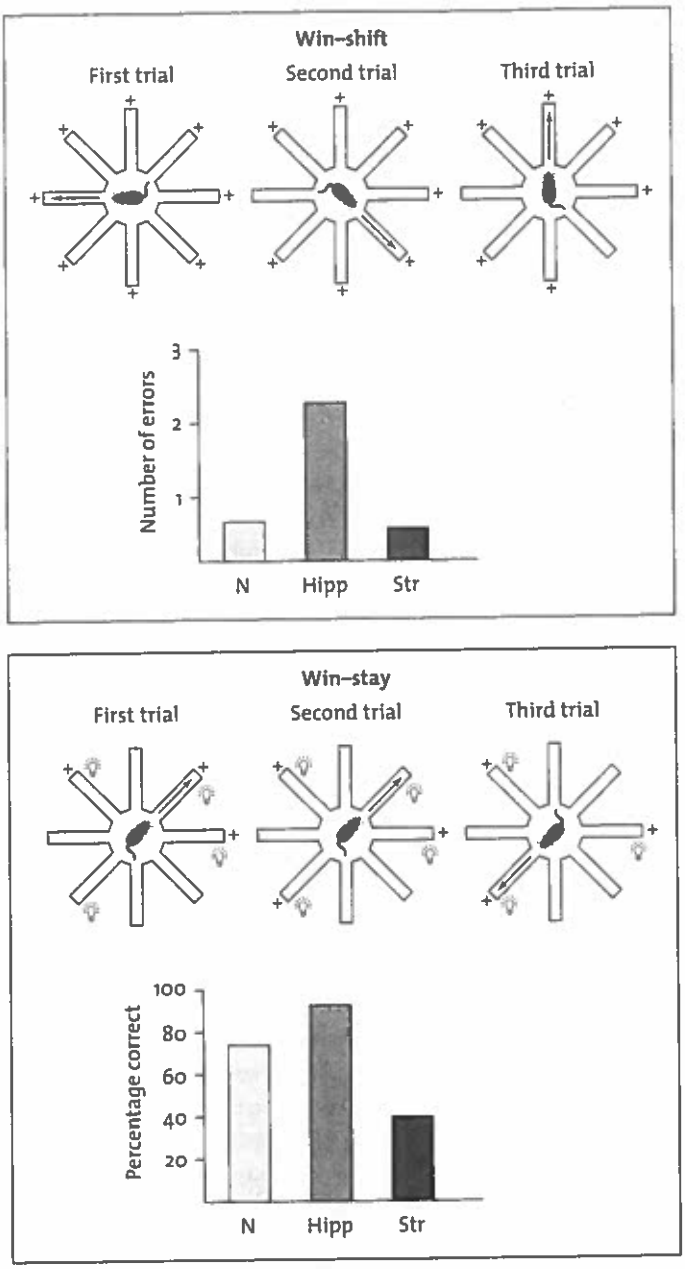

FIGURE 6.6 Hippocampus and striatum lesions and radial maze performance. In the radial maze diagrams, + indicates an arm that was baited with food, and a light bulb indicates an arm that was illuminated. N = normal control subjects; Hipp = subjects with damage to the hippocampus; Str = subjects with damage to the striatum.

though this brought them to a different place. This *response memory* depended on the striatum, not on the hippocampus. These findings offer compelling evidence of distinct learning strategies mediated by the striatum and hippocampus and show that neither brain area is required for the perceptual, motivational, or motor coordination demands common to both response and place learning.

Learning & Memory in Action

Why Does Stress Often Cause Forgetting?

When some people are stressed, they say, their thinking shuts off and they struggle through the situation on a sort of autopilot, remembering few details while learning how to react when the same situation recurs. This anecdotal observation suggests that emotional stress might somehow shift the balance of memory systems from the predominant use of a hippocampus-dependent cognitive strategy to the predominant use of a striatum-dependent habit strategy.

To investigate this possibility and identify its neural basis, Packard and Wingard (2004) examined the effects of stress on place and response learning in a maze. They treated rats with a drug that produces anxiety before giving them a maze task. After the drug wore off, the rats were tested to see which strategy they used. Rats treated with a placebo predominantly used a place strategy; those treated with the anxiety-creating drug predominantly used a response strategy. In a second experiment Packard and Wingard infused small quantities of an anxiety-creating drug directly into each rat's amygdala, which sends emotional signals to the striatum and hippocampus (see Figure 6.5). These findings matched those of the general treatment with the same drug. Thus anxiety and stress appear to make the amygdala switch off the hippocampus so that the striatum becomes the predominant learning system.

The Striatum Also Supports Human Instrumental Learning

Our understanding of the striatum's role in human learning and memory comes mainly from studies of individuals with Parkinson's disease or Huntington's disease. Both disorders lead to profound motor deficits. Parkinson's disease causes tremors, rigidity, and akinesia (inability to move) following cell death in the substantial nigra. This cell death depletes the neurotransmitter dopamine in the striatum. Huntington's disease, which is also characterized by lost striatal function, causes irregular movement patterns. As you might imagine, studying people with these diseases to characterize the role of the striatum in human memory is far

more complicated than examining animals' brains. In addition to their motor deficits, some people with Parkinson's disease also suffer from depression or dementia; people with Huntington's disease always develop progressive dementia. Furthermore, the drugs used to treat these symptoms can affect cognition. These factors can unintentionally confound or influence experimental results, making them more difficult to interpret.

Despite these limitations, clinical observations of striatum function have demonstrated that people with Parkinson's disease or Huntington's disease show deficits in several skill learning tasks (Salmon & Butters, 1995). In one study, people with Huntington's disease were trained in rotary pursuit (Gabrieli, 1995)—a simple motor skill learning task that requires subjects to maintain contact between a handheld stylus and a metal disk revolving on a turntable. Normal subjects improved with practice, increasing the amount of time they maintained contact with the target; but subjects with Huntington's disease showed virtually no learning. Because Huntington's disease causes specific motor deficits that might make the rotary pursuit task harder, the turntable was slowed so that when the Huntington's subjects were introduced to the task, their initial performance was as good as that of control subjects working with a faster rotating disk. Equating initial performance levels in this way did not reduce the learning deficit; even when the task was adjusted so that the Huntington's subjects' initial performance exceeded that of normal subjects, the Huntington's subjects still failed to show learning. Thus the researchers showed that the striatum contributes to motor skill learning.

A more complex skill learning task in which individuals with Huntington's disease and Parkinson's disease are impaired is the serial reaction time motor sequencing task. In this task, a computer screen shows four lighted panels, each corresponding to a button on the keyboard below. One screen location flashes during each trial, and the subject presses the button corresponding to that location. Unknown to the subject, the locations flash in a repeating order (such as a 12-item sequence in which each of the four locations is flashed three times). Implicit learning of the sequence is measured in two ways. First, as subjects learn the sequence, their average reaction time to respond to a given item gradually decreases. Second, reaction time slows substantially in a test period when the stimuli are presented in random order. The shorter reaction times for the learned series compared to those for a random series reflect learned memory. Subjects with Parkinson's disease or Huntington's disease are typically impaired in motor sequence learning (Willingham & Koroshetz, 1993; Pascual-Leone et al., 1993). Thus the striatum is vital for learning and remembering the order of movements in a skilled routine.

The striatum is also involved in learning S–R associations in humans, just as it is in animals. A striking example of a double dissociation between the roles of the human striatum and hippocampus involves the analysis of an unusual form of

habit learning. Knowlton and colleagues (1996) compared individuals in the early stages of Parkinson's disease, who had suffered striatum damage, to amnesic individuals with damage to the hippocampus or the hippocampal memory system. These subjects were trained in probabilistic classification learning, presented in the form of a weather prediction game. The task involved using cues from a set of cards to predict one of two outcomes (rain or sunshine). On each trial, one or more cards from a deck of four were presented. Each card was associated with the sunshine outcome only in a probabilistic way (Figure 6.7). For example, a particular card predicted sunshine in 60 percent of the trials, and a different card predicted sunshine in 40 percent of the trials. The stimuli differed only in the pattern of shapes printed on the cards, and the subjects were not told the probability of sunshine associated with each card. When multiple cards were presented, the correct weather prediction was determined by the average of their probabilities. For example, if one card of a presented pair predicted sunshine in 60 percent of the trials and the other card predicted sunshine in 40 percent of the trials, the average probability was 50 percent. Based on the cards presented in each trial, the subject had to predict rain or sunshine and was then given feedback about the prediction's accuracy.

The probabilistic nature of the task made it counterproductive for subjects to attempt to recall earlier trials because any particular cue configuration could lead to more than one outcome. For example, when presented with the pattern shown in a previous trial, a subject might remember the response she made for the earlier occasion; but the current trial's outcome need not be the same as in that earlier trial, leading to confusion. Instead the most useful information to be learned concerned the probability associated with particular cues and cue combinations, so that learning occurred gradually across trials—much as real-life habits or skills are acquired. This format was specifically designed to eliminate any advantage normal subjects would have in remembering specific prior trials.

Over the initial trials, normal subjects gradually improved from pure guessing (50 percent correct) to about 70 percent correct—about as accurate as possible because the outcomes were probabilistic. However, subjects with Parkinson's disease failed to show significant learning, and this impairment was particularly evident in those with more severe Parkinsonian symptoms. By contrast, the amnesic subjects learned the task, achieving accuracy levels like those of controls by the end of the trials.

After this weather prediction training, the subjects were asked multiple-choice questions about the types of stimulus materials and nature of the task. Normal subjects and those with Parkinson's disease performed well in recalling the task events. But the amnesic subjects were impaired, performing near the chance level of 25 percent correct. What do these results suggest about the relationship between the kind of memory required to learn a skill and the memories we have of particular events—such as the time spent learning that skill? As you probably noticed,

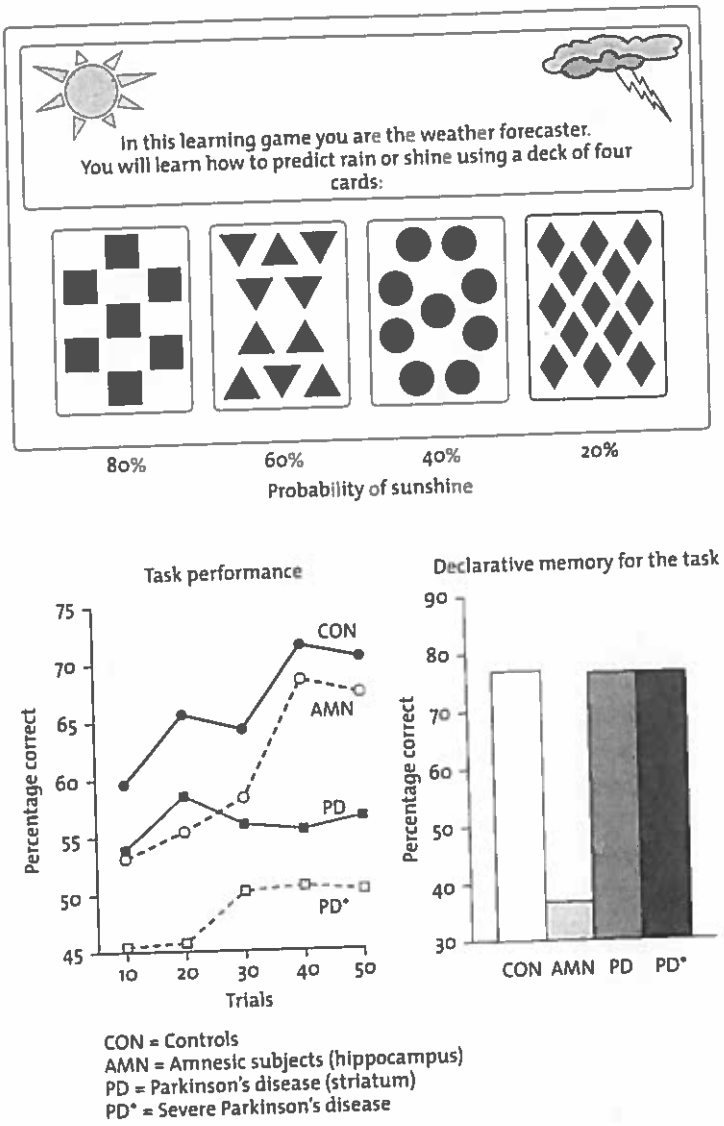

FIGURE 6.7 Weather prediction task.
A. Appearance of the computer screen on a trial using all four cards. B. Performance in forecasting weather (left) and in memory for the task stimuli and rules (right).

these findings demonstrate a clear double dissociation: Habit learning is disrupted by striatum damage, and memory for learning events is impaired in amnesia—providing further evidence that habit learning depends on the striatum whereas ordinary memory for identical learning materials is accomplished within a sepa-

rate brain system. The amnesic subjects in this study were similar to Jimmie G. (Chapter 1), who learned his way around a hospital garden but professed no memory of having seen it before.

Instrumental Learning Is Involved in Drug Abuse Chronic abuse of addictive drugs is a complex social phenomenon based partly in learning and brain mechanisms. Two features of addictive behavior make drug abuse particularly problematic. First, addiction is compulsive: Addicts are driven to abuse even when they are aware of its deleterious affects on their lives. Second, addiction is subject to relapse even after seemingly successful withdrawal and periods of abstinence. Substantial recent evidence suggests that S–R learning and its brain system play a central role in these two prominent aspects of drug addiction. Different abused drugs, including amphetamine, heroin, nicotine, and alcohol, have distinct molecular mechanisms; but all increase the release of dopamine in the striatum. This fact has led to the suggestion that the reinforcing effects of drugs may be mediated by the same brain systems involved in learning based on natural rewards (Everitt et al., 2001; Berke & Hyman, 2000; Kelley, 2004). By altering the release of dopamine in the striatum, drugs could change information processing in the striatum and throughout the brain system that mediates S–R learning. Specifically, to the extent that dopamine is a reinforcement signal in the striatum, drugs may achieve some of their dramatic effects by enhancing the operation of the S–R learning system. In addition, the neurotransmitter glutamate and NMDA receptors play a crucial role in plasticity within the circuits identified in Figure 6.5.

Within the frameworks proposed, we typically employ our capacity for foresight and conscious control of long-term consequences to adjust our reward-seeking behavior. However, current theorizing suggests that the effects of drug abuse on striatal dopamine may enhance S–R learning within that system to overwhelm mechanisms of cognitive control, leading to both compulsive behavior and relapse. Consistent with this view, drug-seeking behavior of addicts becomes increasingly stereotyped and ritualized, automatic, inflexible, and stimulus-bound. In particular, self-reports of relapse suggest that addicts lose conscious direction in reacting directly to environmental cues that take control of behavior. Indeed, cravings for addictive drugs depend greatly on contextual cues associated with drug taking. In a hospital setting, cravings usually diminish—but they can become acute as soon as the addict reenters the environment in which drugs were a routine of life. For example, entering a bar is too strong a stimulus for alcoholics, and returning to the social milieu of drug taking often incites rapid relapse in heroin addicts. Similarly, in animal models of drug addiction, following extinction of drug taking, reintroducing environmental cues associated with former addictive behavior provides powerful cues for reinstatement of drug taking. Considerable research on dopamine, NMDA receptors, and striatal system function is providing a possible new avenue of success in treating addiction.

The Striatum Is Activated during Habit Learning in Humans

Functional neuroimaging studies provide another way to explore the role of the striatum in human learning and memory, with results closely paralleling studies of people with brain damage. In these studies, subjects perform an instrumental learning task while their brain activity is measured by blood flow changes in the striatum and other brain areas. Increases in striatal activation have been associated with learning finger movement sequences (Seitz et al., 1990; Seitz & Roland, 1992) and with learning to press buttons following a sequence of panel lights on a computer screen (Grafton et al., 1995). Neuroimaging studies have also documented striatal activation during perceptual and cognitively based skill learning tasks like probabilistic classification in the weather prediction task (Poldrack et al., 1999). Taken together, the neuropsychological results and neuroimaging evidence show that the role of the striatum in habit or skill learning encompasses a variety of learning abilities that involve multiple stimuli and response choices and that show gradual, incremental performance improvement across trials.

Striatal Neurons Are Activated during Habit Learning

Recent neurophysiological studies have confirmed the involvement of the striatum in S–R learning and have revealed how the striatum represents memories. Several studies have described neurons in the striatum that fire in anticipation of movements, suggesting that striatal activity might be associated with the relationship between behavioral contexts and responses (Mink, 1996). Also, neurons that project from the brain stem into the striatum and use dopamine as a neurotransmitter fire with the expectation and reception of rewards (Schultz et al., 1993; Schultz, 2006). Moreover, these dopamine neurons carry signals to the striatum about the predictability of rewards contingent on a stimulus. Recall from Chapter 5 that Rescorla and Wagner (1972) showed that conditioning occurs only when a stimulus adds predictive value regarding a reinforcer. An added stimulus that does not alter predictions is not learned, as shown in the phenomenon of blocking. Consistent with these behavioral findings, once a contingency between a stimulus and reward is established, adding a nonpredictive stimulus does not activate the striatal dopamine neurons, whereas these neurons do respond to a stimulus that adds predictive value (Figure 6.8). The combined findings for neurons in the striatum and elsewhere in this circuit have led researchers to suggest that the striatum uses knowledge about the behavioral context and reward predictions to plan behavioral responses (Schultz et al., 2000; Graybiel et al., 1994).

In addition to these observations about simple and conditional motor responses, other data indicate a prominent role for the striatum during sequence learning.

Kermadi and Joseph (1995) trained monkeys to fixate on a central location and memorize a sequence of panels that were illuminated in order on a computer screen, then look at and subsequently reach toward each panel in the correct order (Figure 6.9). Many striatal neurons responded to the visual instruction stimuli while the monkeys fixated on the initial point or during their eye or arm movements. Furthermore, the responses of many neurons depended on the sequential order of the targets: They responded to a visual cue only if it was in a particular position in the three-item sequence. Notably, these cells fired in anticipation of each item in the sequence, demonstrating again that the striatum participates with other structures in anticipating behavior sequences.

These neurophysiological observations converge with anatomical and behavioral data to suggest that the striatum plays a critical role in habit learning by resolving competition among multiple sensory and response options, particularly as we learn response sequences. The striatum houses circuitry for cortical sensory input and direct motor outputs, which are both necessary to associate stimuli with specific behaviors. Furthermore, the striatum receives reward signals capable of reinforcing associations of stimuli and responses. So the striatum is a key component of a pathway for habit learning involving the acquisition of stereotypes and unconscious behavioral patterns, and this pathway can mediate S–R behavior independent of circuits for other forms of memory.

Interim Summary

In both animals and humans, the striatum plays a central role in habit and skill learning. Animal experiments have shown double dissociations between the effects of damage to the striatum versus the hippocampus. Striatal damage causes deficits in S–R learning in which a specific cue (such as a light) should be approached or a particular response (like a left turn) should be executed to receive reinforcement; however, striatal damage does not harm memory of a reward's location or the reward value expected. By contrast, the hippocampus is not essential for S–R learning but instead supports memory for where rewards were obtained.

Humans with Parkinson's and Huntington's diseases have striatal dysfunction, resulting in motor coordination deficits and S–R learning impairments in both simple motor skill learning tasks and complex sequence learning tasks. Complementary studies of striatum physiology show that the striatum is activated when humans learn finger movement sequences, as well as during S–R learning in a probabilistic classification task. In animals performing S–R and sequencing tasks, striatal neurons are activated when the animals anticipate movements or rewards; these neurons also fire in association with learned movement sequences.

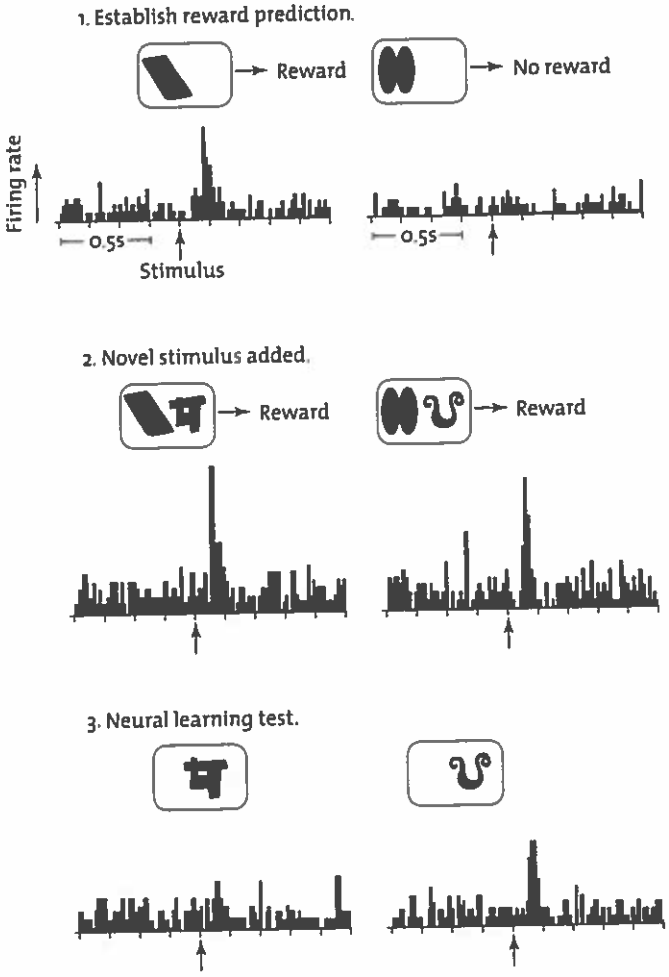

FIGURE 6.8 Blocking of neural predictive responses in a dopamine neuron. The top panels show that this cell fires after presentation of a stimulus that predicts reward and does not fire after a different stimulus that predicts no reward. The middle and bottom panels show that adding a second stimulus to the reward-predicting stimulus does not further activate the cell; but adding a second stimulus that changes the reward prediction of the formerly negative stimulus activates the cell.

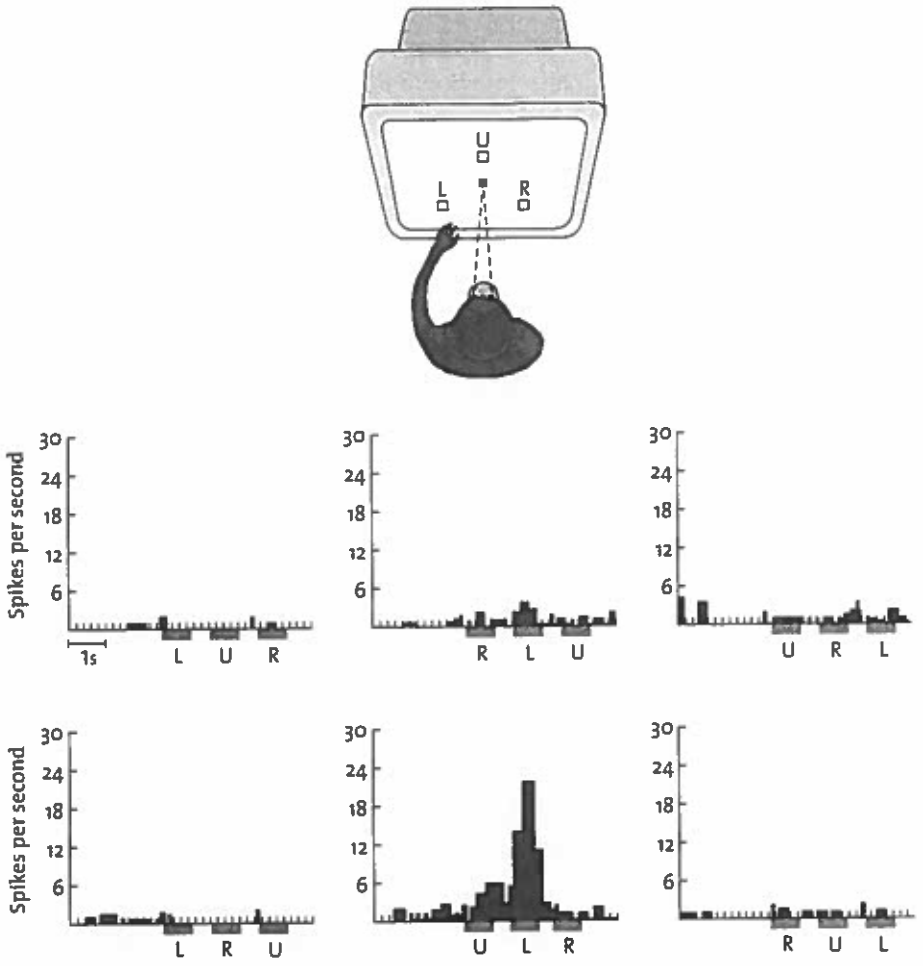

FIGURE 6.9 Striatal neuron firing patterns in the order task.
R, U, and L refer to stimulus positions in the right, upper, and left, respectively. Note the activation of this neuron selectively during the middle part of the ULR sequence.

Chapter Summary

1. Instrumental learning lets us acquire specific reinforced behavioral responses. Habits are based on predictable consequences of actions. Habit acquisition pervades studies of learning and memory using mazes and a variety of operant conditioning tasks in animals; it is also prominent in human skill learning and implicit learning of specific responses to stimuli and response sequences.

2. Unlike classical conditioning, in which the US is delivered regardless of the response, in instrumental learning the reinforcer is provided only when a particular behavior is emitted. As embodied in Thorndike's Law of Effect, behaviors associated with positive reinforcement increase in likelihood, whereas behaviors associated with negative reinforcement decrease in likelihood. Many superstitious behaviors increase in occurrence when associated with reinforcement; but only some behaviors (called terminal responses) that occur near the time of reinforcement obey the Law of Effect. Behaviors must be consistent with their reinforcers—that is, they must be related or "belong" to the rewards—to be learned. In instrumental conditioning, animals learn associations between a stimulus, a response, and a reinforcer and can act distinctly for each such association; animals behave according to their reward expectancy for a given stimulus and response.

3. Various mazes have been used to study instrumental learning. Rats can learn to return to maze locations consistently associated with rewards (win–stay), and they can remember specific locations and not return to already visited places (win–shift). Another prominent example of instrumental learning is the free operant task, in which rats and pigeons can emit specific responses at any time and are reinforced according to ratio or interval schedules. Animals can learn reward schedules, match their response distributions to the reward probabilities of multiple response choices, learn to chain operant responses, and learn serial patterns. Humans also learn sequences of behaviors called skills. Skilled performance is acquired in stages and is mediated by motor programs that contain the sequence of coordinated movements.

4. Instrumental and skill learning are supported by a brain system that involves cortex and striatum circuitry. Animal and human studies have shown that instrumental learning of reinforced responses depends on the striatum. Many of these studies have distinguished striatum-dependent habit learning from hippocampus-dependent memory of specific places or events. Correspondingly, striatum neurons are activated during instrumental and skill learning, particularly that involving learning behavior sequences.

5. Chronic drug abuse shares many features with the phenomenology and circuitry of the instrumental learning system. Addictive conduct can be character-

ized as an exaggerated instrumental behavior that dominates conscious control of the consequences of actions. Instrumental conditioning is also relevant to the shaping of behaviors in many clinical and workplace situations.

KEY TERMS

instrumental conditioning (p. 183)
habit (p. 184)
reinforcers (p. 184)
Law of Effect (p. 185)
superstitious learning (p. 186)
terminal responses (p. 187)
interim responses (p. 187)
belongingness (p. 187)
reinforcer devaluation (p. 188)
learned helplessness (p. 189)
free operant (p. 193)
Skinner box (p. 194)
shaping (p. 194)
secondary reinforcer (p. 194)
autoshaping (p. 195)
behavioral modification (p. 196)
token economy (p. 196)
partial reinforcement effect (p. 197)
fixed ratio schedule (FR) (p. 197)
variable ratio schedule (VR) (p. 197)
fixed interval schedule (FI) (p. 197)
variable interval schedule (VI) (p. 197)
cumulative response (p. 197)
matching law (p. 198)
probability matching (p. 199)
response chaining (p. 199)
backward chaining (p. 200)
serial patterns (p. 200)
motor program (p. 202)
cognitive stage (p. 202)
associative stage (p. 202)
autonomous stage (p. 203)
double dissociation (p. 204)

REVIEWING THE CONCEPTS

- How does instrumental learning differ from classical conditioning? What associations are learned in instrumental learning?
- What is the Law of Effect, and what role does it play in instrumental learning? Are all behaviors equally subject to the Law of Effect?
- How do superstitious behaviors appear? How can we get rid of them?
- How do rats solve maze problems?
- How is behavior affected by reinforcement schedules?
- How are response chains acquired?
- What are the psychological stages of skill learning in humans?
- Which neural pathways support habit and skill learning?
- How are instrumental conditioning mechanisms involved in drug abuse?