From 1977 to 1992, the country of Mozambique was caught in a brutal civil war. When the conflict ended, tens of thousands of land mines were left buried over millions of square meters. In the decades that followed, several hundred civilians were seriously injured or killed each year while attempting to cross or farm the mined land. For many years, local volunteers and international aid workers shared the dangerous task of going meter-by-meter through the minefields, clearing them of bombs.

Along with traditional approaches such as the use of metal detectors to find buried explosives, one enterprising organization trained giant African pouched rats to detect the scent of the bombs and signal to their handlers that a mine was nearby. Because the rats weigh only about 1 to 2 kilograms, they can scamper across minefields safely, without detonating the devices. One trained rat can check an area of 200 square meters in about 30 minutes—a job that might take several days for a human using a metal detector, partly because the device would be distracted by stray pieces of scrap metal, which the rats ignore (Kristof, 2015). In the span of one year (2009), a team of 34 rats and 50 handlers cleared nearly 200,000 square meters, finding 75 land mines (Poling et al., 2010). By 2015, Mozambique officially declared that it was free of land mines. The rats are still being used in nations such as Angola and Cambodia, where other conflicts have left in their wake numerous unexploded land mines and other buried ordnance.

How do these mine-detecting rats learn their jobs? The training is intensive, and it typically takes 9 months for a rat to be “accredited” and deployed in the field, but the basic idea is simple. The rats learn that, when they detect a target scent (here, the odor of explosives) and make a specific behavioral response (such as pawing at the ground) to alert the handler, they may earn a small reward (such as a piece of banana). The rats like the treats, so they willingly scamper across fields to seek out the target smell and signal its presence in order to earn more rewards.

This kind of learning is an example of operant conditioning, the process whereby organisms learn to make or to refrain from making certain
The process whereby organisms learn to make or to refrain from making certain responses in order to obtain or avoid certain outcomes. The learning is called “operant” because the organism “operates” on the environment in a way that causes an outcome to occur. Operant conditioning is sometimes also called instrumental conditioning, meaning that the organism’s behavior is “instrumental” in producing the outcome.

Operant conditioning sounds deceptively simple, but the real-world applications are complex and profound. In addition to mine detection, African pouched rats are also being trained to sniff sputum samples from patients for evidence of tuberculosis in parts of the world where locals cannot easily afford expensive laboratory testing (Mgode, Cox, Mwimanzi, & Mulder, 2018). Closer to home, you may frequently see service dogs that help visually impaired human companions navigate busy city streets or sniffer dogs that search for drugs and explosives at airports and border crossings. One innovative group of researchers has even trained honeybees to signal when they detect the scent of illicit drugs (Schott, Klein, & Vicleinskas, 2015), raising the possibility that at some point in the future, police officers on routine traffic stops might carry along a box of bees. By the end of this chapter, you should have a fairly good idea of how all these animals learned their trades.

5.1 Behavioral Processes

Humans have been using the principles of operant conditioning as long as there have been sheep to herd, horses to ride, and puppies to housebreak. And parents throughout history have used operant conditioning when praising or rewarding their children for desired behavior and when scolding or punishing their children for misbehavior. But it was not until the end of the nineteenth century (about the same time that Ivan Pavlov was “discovering” classical conditioning) that Edward Thorndike first tried to systematically explore how animals learn new behaviors.

The “Discovery” of Operant Conditioning

Chapter 1 describes Edward Thorndike and his studies of how cats learned to escape from puzzle boxes, such as the one shown in Figure 5.1A. The puzzle boxes were made from fruit crates, each with a door that could be opened from the inside if the animal executed the correct sequence of pressing levers, pulling ropes, and stepping on pedals (Thorndike, 1898, 1911, 1932). When Thorndike put a cat in such a box for the first time, the animal would hiss, claw at the walls, and generally try to fight its way free. Eventually, the cat would accidentally perform the movements needed to open the door and get out. Thorndike recorded how long it took the animal to escape and then returned the cat to the box to try again. After a few experiences in the box, a cat would typically learn an efficient sequence of moves that allowed it to escape. Figure 5.1B shows the data from one cat; after a dozen or so trials in the box, the cat was able to get out almost immediately.

Thorndike concluded that when an animal’s response was followed by a satisfying outcome (such as escaping from a puzzle box or obtaining food), the probability of that response occurring again in the future would increase. Thorndike formalized this idea in his law of effect, which is presented in
Chapter 1. Specifically, in the presence of a particular stimulus, called the **discriminative stimulus** ($S^D$), a particular response ($R$) may lead to a particular outcome ($O$).

$S^D$ is called a “discriminative stimulus” to emphasize that it helps the organism “discriminate” the conditions under which $R$ will lead to $O$. If the outcome $O$ is desirable or pleasant, the response $R$ will tend to increase in frequency, strengthening the $S^D \rightarrow R$ association and making it more likely that $S^D$ will evoke the same $R$ in the future.

But as Chapter 1 also describes, Edward Tolman argued that Thorndike’s $S^D \rightarrow R$ framework was too limiting; Tolman believed that animals make responses because they (in some sense) understand that response $R$ leads to a specific, predicted outcome $O$. More recent studies have supported many of Tolman’s ideas and have even identified places in the brain where $S^D \rightarrow R$ associations may be stored and where expected outcomes $O$ are processed (more on that later, in Section 5.2). For now, we will formalize operant conditioning as a three-part association:

Discriminative stimulus $S^D \rightarrow$ Response $R \rightarrow$ Outcome $O$

In the case of a puzzle box, $S^D$ is the box, $R$ is the sequence of movements needed to open the door, and $O$ is the escape. The $S^D \rightarrow R$ association is strengthened when $R$ is followed by a desirable outcome $O$.

**Classical Versus Operant Conditioning**

A few decades after Thorndike published his original account of operant conditioning, B. F. Skinner read it, along with Pavlov’s work on classical conditioning, and concluded that these two types of learning are fundamentally different (Skinner, 1938). In classical conditioning (described in Chapter 4), organisms experience an outcome (the unconditioned stimulus, or US) whether or not they perform the conditioned response (CR). But in operant conditioning, the outcome $O$ depends on whether the organism performs the response $R$. 

**discriminative stimulus ($S^D$)**
In operant conditioning, a stimulus indicating that a particular response ($R$) may lead to a particular outcome ($O$).
For example, in classical conditioning of the eyeblink response, a rabbit may hear a tone (the conditioned stimulus, or CS) that is reliably followed by an airpuff US, and the rabbit may learn to make an eyeblink CR to the tone. The airpuff follows the tone whether or not the CR occurs—so this paradigm is classical conditioning. By contrast, a cat placed in a puzzle box (S^D) must learn to make a series of responses (R) in order to escape and obtain food (O). If the responses R are not performed, the outcome O does not occur. Therefore, this paradigm is operant conditioning. When deciding whether a paradigm is operant or classical, focus on the outcome. If the outcome occurs regardless of responding, the paradigm is classical; if the outcome is contingent on a response, the paradigm is operant.

Differences aside, operant and classical conditioning share many characteristics, including a negatively accelerated learning curve: Figure 5.1B shows that the time to escape from a puzzle box decreases rapidly in the first few trials and then levels off; similar leveling off occurs in classical conditioning (as seen in Figure 4.8). Both operant and classical conditioning also show extinction: a tendency for learned responses to extinguish if no longer followed by the outcome. (Extinction is discussed in more detail in Chapter 4.)

**TEST YOUR KNOWLEDGE**

**Is It Classical or Operant?**

In classical conditioning, the outcome (US) follows the stimulus (CS) whether or not a learned response (CR) is performed. In operant conditioning, the outcome (O) follows the discriminative stimulus (SD) only if a particular response (R) is performed. Analyze the following scenarios to check whether you understand the difference. (*Answers appear at the end of the chapter.*)

1. Since retiring, Jim spends a lot of time sitting on his back porch, watching the birds and whistling. One day, he scatters crumbs, and birds come and eat them. The next day, he sits and whistles and strews crumbs, and the birds return. After a few days, as soon as Jim sits outside and starts whistling, the birds arrive.

2. Shevonne’s dog Snoopy is afraid of thunder. Snoopy has learned that lightning always precedes thunder, so whenever Snoopy sees lightning, he runs and hides under the bed.

3. Michael has accepted a new job close to home, and now he can walk to work. On the first morning, there are clouds in the sky. It starts to rain while Michael is walking to work, and he gets very wet. The next morning, there are again clouds in the sky. Michael brings his umbrella along, just in case. When it rains, he stays dry. After that, Michael carries his umbrella to work anytime the sky looks cloudy.

4. In Carlos’s apartment building, whenever someone flushes the toilet, the shower water becomes scalding hot, causing him to flinch. Now, whenever he’s in the shower and hears the noise of flushing, he automatically flinches, knowing he’s about to feel the hot water.

**Free-Operant Learning**

Chapter 1 introduces B. F. Skinner, the “radical behaviorist.” Skinner was attracted to Thorndike’s work, with its promise of animal responses that could be measured and evaluated without requiring speculation about the animal’s mental states. But Skinner thought he could refine Thorndike’s techniques. Thorndike’s procedures were characterized by *discrete trials*, meaning that the experimenter defined the beginning and end of
each trial. For example, on one trial with the puzzle box, the experimenter would pick up a cat, put it in the box, shut the door, and record how long it took the cat to escape. He would then pick up the cat and return it to the box to start the next trial. Thorndike also built mazes (constructed out of books placed on end in various configurations) and measured how long it took animals (usually baby chicks) to find their way out. On each trial, he would place a chick at the start of the maze and record how long the chick wandered around the maze before it found its way out. He would then pick up the chick and return it to the start to begin the next trial. This type of experimental setup is often referred to as a discrete-trials paradigm because each trial is separate, or discrete, and the experimenter decides when and how often to begin a new trial.

Skinner developed a maze with a return ramp so that the rat could finish one trial in the maze, collect the food, and then run back to the beginning of the maze by itself to start the next trial—and obtain the next piece of food. Part of Skinner’s intention was to automate data collection so that the experimenter no longer had to intervene at the end of each trial by returning the rat to the starting position for the next trial. But a side effect was that the animal—not the experimenter—now controlled its own rate of responding, by how quickly or slowly it ran around to start the next trial. This type of setup is often referred to as a free-operant paradigm, meaning that the animal can operate the apparatus freely, whenever it chooses.

To measure behavior more directly, Skinner also devised a cage—now commonly called a Skinner box—within which food could be delivered automatically (Figure 5.2A). The box contained a mechanism, such as a lever or a pressure-sensitive disk, that controlled the delivery of food. When the animal pressed the lever or tapped the disk, food dropped into the trough. As the animal explored its cage, eventually it would accidentally manipulate the lever or disk and receive the food. Over time, as animals learned the relationship between the response R (pressing the lever or disk) and the outcome O (obtaining food), they would dramatically increase their rate of responding.

The experiment can be made a bit more elaborate by adding a discriminative stimulus SD, such as a light on the cage wall, that signals whether response R will be reinforced; for example, pressing the lever while the light is on will trigger food delivery (outcome O), whereas lever presses while the light is off will not. Gradually, the animal learns that, in the presence of the light (SD), lever presses (R) will result in food (O); as a result, the frequency of R when SD is present will increase.

Figure 5.2B shows an example of the data that might be obtained from a free-operant experiment in which lever presses (R) in the presence of the light (SD) are reinforced by food delivery (O) for the first 13 minutes of the experiment. During this acquisition period, the number of lever presses per minute increases rapidly, as the animal learns the SD → R → O association. This acquisition period is followed, in minutes 14 to 26 of the experiment, by an extinction period, meaning that the outcome (here, food delivery) no longer occurs no matter what response the animal makes. During the extinction period, the frequency of R in the presence of SD gradually decreases to near zero, as the animal learns the new association SD → R → no O.

Skinner next invented a means of recording responses automatically. Back before the advent of modern computers, mechanical devices such as the one shown in Figure 5.2C recorded data on a long piece of paper rolling steadily beneath a pen. (Until fairly recently, such devices were also used for seismographs and lie detectors.)
FIGURE 5.2 Operant conditioning

(A) A Skinner box in which lever-press responses are reinforced by delivery of food into a food cup. (B) Hypothetical data illustrating learning by a rat in a Skinner box, shown as the mean response rate during a 26-minute experiment. During the first 13 minutes (acquisition phase), lever presses are reinforced by food delivery, so the rate of responses per minute increases; by the final (thirteenth) minute, the rat is lever-pressing almost 10 times per minute. During the last 13 minutes (extinction phase), lever presses are no longer reinforced by food delivery, so the rate of responding decreases until the rat is making 1 or fewer responses per minute. (C) A cumulative recorder provides another way to display the data from part B. Here, a paper scrolls steadily leftward, causing a pen to draw a flat line across the paper. Each time the animal makes a response, the pen moves up to a new position, jittering the line up slightly. (D) The output from a cumulative recorder, plotting the data from B. Here, the steep upward slope during the first 13 minutes (acquisition) reflects the fact that the rat is responding more and more often; the flattened slope in the final 13 minutes (extinction) shows the rate of responding is petering out. Graphs in B and D both show the same data, plotted in different ways.

Skinner hooked up one of these devices to a Skinner box so that the pen would move up slightly to a new, higher position each time the animal responded. If the animal made no responses, the pen drew a long straight line as the paper scrolled by. But whenever the animal made a response, the pen ticked up. If the responses came quickly enough, the resulting line sloped upward steeply (Figure 5.2D). The device shown in Figure 5.2C is called a cumulative recorder because the height of the line at a given time shows the total number of responses made up to a given time.

cumulative recorder
A device used for recording responses in operant conditioning, designed in such a way that the height of the line it draws represents the total (cumulative) number of responses made up to a given time.
One modern example of a cumulative recorder is the odometer in a car. The odometer ticks off miles driven, and the ticks occur faster if you drive faster. When you park the car for the night and then start it up the next morning, the new mileage is added right on top of the old, for a cumulative record of total miles driven so far.

Although pen-and-paper cumulative recorders are not in general use anymore, data from operant conditioning experiments are still sometimes reported as cumulative responses (as in Figure 5.2D) instead of as response rates (as in Figure 5.2B). The label on the y-axis will tell you which kind of data you are looking at. The actual behavior being recorded is the same, whichever type of graph is used.

**Learning About Stimuli and Responses**

We have now defined operant conditioning as a three-part association between $S^D$, $R$, and $O$. But, in fact, each of these components ($S^D$, $R$, and $O$) can independently influence what is learned.

**Discriminative Stimuli**

Swimmers line up at the edge of a pool before a race. At the sound of the starting whistle, they dive in as quickly as possible to get a good start in the race. But any swimmer who dives in too early (before the starting whistle) may be penalized or even disqualified. The outcome varies, depending on whether the dive occurs before or after the whistle.

Discriminative stimuli help the learner identify (or “discriminate”) the conditions in which a response will be followed by a particular outcome. For the swimmers, the starting whistle is a discriminative stimulus signaling that dive responses will now result in a favorable outcome:

$S^D$ (starting whistle) $\rightarrow$ $R$ (dive) $\rightarrow$ $O$ (good start in the race)

In a Skinner box, a light may be the discriminative stimulus. This means that lever-press responses while the light is on result in the outcome of food delivery, but lever presses while the light is off do not:

$S^D$ (light on) $\rightarrow$ $R$ (press lever) $\rightarrow$ $O$ (get food)

$S^D$ (light off) $\rightarrow$ $R$ (press lever) $\rightarrow$ $O$ (no food)

As the $S^D$ $\rightarrow$ $R$ $\rightarrow$ $O$ notation suggests, the discriminative stimulus is the first part of the chain that triggers the response and leads to the outcome. Sometimes, the $S^D$ $\rightarrow$ $R$ association is so strong that the discriminative stimulus $S^D$ seems to evoke the learned response $R$ automatically, no matter what other options are available. In one striking example, well-trained rats in a familiar maze ran right through a pile of food on their way to the goal box (Stoltz & Lott, 1964). Apparently, the discriminative stimulus of the maze environment $S^D$ was so strongly associated with the maze-running response $R$ that unexpected food encountered along the way could not disrupt the $S^D$ $\rightarrow$ $R$ association. Such behavior is sometimes called a *habit slip*. People exhibit habit slips all the time, particularly when drowsy or distracted. Perhaps you have started driving to a friend’s house, only to find that your attention lapsed, and the car is now heading along the more frequently traveled route to school; or perhaps you have awoken late one morning and started hurriedly dressing for class, only to realize today is Saturday, and you can stay in bed awhile. If so, you have experienced the effects of a strong $S^D$ $\rightarrow$ $R$ association.
Responses

In operant conditioning, the organism learns to make a specific response R that produces a particular outcome O. A response is defined not by a particular pattern of motor actions but rather by the outcome it produces. For example, a rat in a Skinner box may receive access to food when it presses a lever:

\[ S^D (\text{lever in box}) \rightarrow R (\text{press lever}) \rightarrow O (\text{get food}) \]

If the lever is pressed, the food arrives—whether the rat presses the lever with its left front paw, its right front paw, or even its nose. Anything the rat does to press the lever sufficiently will trigger the food delivery device. In one experiment by American psychologist Karl Lashley, rats were trained to run through shallow water in a maze, making a correct series of turns to reach the goal; later, Lashley raised the water level so that the animals had to swim. The rats continued to navigate to the goal, even though swimming involved a new set of motor responses that the rats had never executed in this environment (Lashley, 1924). Similarly, parents may reinforce a desired behavior—say, neatness—by providing an allowance to their daughter Becky if she tidies her room; the precise sequence of movements by which Becky accomplishes this task is unimportant, as long as the clothes come off the floor and the toys are put away.

But in cases this complex, how do children and animals learn which responses lead to particular outcomes in the first place? In the case of Thorndike’s cats and chicks, learning occurred by trial and error; on the first trial, each animal could take minutes or hours before happening to execute the sequence of responses that opened the puzzle box door or led out of the maze. Now imagine training a service dog to help a blind man cross a busy city street. This requires that the dog approach the curb, stop, wait for a green traffic light, check for oncoming cars, and then proceed to the far curb while staying within the white painted lines. It is highly unlikely that an untrained dog would ever happen to make that complete sequence of responses and accidentally discover that the trainer would reinforce exactly those behaviors.

As you may have guessed, researchers and animal trainers who want to train complex behaviors rarely rely on accidents. Instead, they have developed special operant conditioning approaches, two of which are described next.

Shaping

A training process called shaping works by reinforcing successive approximations to the desired response. Consider the following example. When a rat is first placed in a Skinner box, it may perform any of its natural behaviors: grooming, exploring, or just sitting quietly. When the rat happens to wander near the food tray, the experimenter drops in a piece of food. The rat eats the food and starts to learn an association between the tray and food. After a few such trials, the rat starts spending time near the food tray. The experimenter then changes the rules so that, now, the rat must also be near the lever before food is dropped. Soon, the rat learns to loiter in the vicinity of the lever. Once the rat has learned this, the rules change again: food is dropped only if the animal is actually touching the lever, then only if the animal is rearing up and touching the lever, then only if the animal is pressing down on the lever. Gradually, by a series of successive approximations, the desired response is learned: the rat presses the lever to obtain food.

If you think this sounds like a difficult and time-consuming process, you’re right. Shaping requires considerable skill on the part of the experimenter, who must decide...
how quickly to proceed, how difficult to make each new stage, and even whether to back up a few steps if the animal appears to be getting confused. Some researchers have proposed standardized methods and criteria that can help optimize shaping techniques (e.g., Galbicka, 1994). Still, the intricacies of shaping are one reason that people often pay professional animal trainers to housebreak puppies.

The principles of shaping can be useful in many situations where difficult learning objectives can be reached by successive approximation. For instance, a physical therapist can use shaping to guide a patient who has suffered a stroke and is attempting to relearn motor control. If the patient has lost the ability to use his left hand, he might first be instructed to practice simply opening and closing his hands before progressively working up to more complex tasks that require fine motor control, such as grasping and using a spoon or holding and writing with a pencil. Shaping has been used to help teach children with autism how to speak, by first reinforcing any vocalizations, then reinforcing only vocalizations that sound like words, and eventually reinforcing actual word production (Lovaas, 1987). Shaping is also used to train service animals, such as guide dogs and sniffer dogs (see “Learning and Memory in Everyday Life,” on page 184, for more on how sniffer dogs are trained and evaluated).

Chaining

A related technique is chaining, in which organisms are gradually trained to execute complicated sequences of discrete responses. Skinner once trained a rat to pull a string that released a marble, then pick up the marble with its forepaws, carry it over to a tube, and drop it inside the tube (Skinner, 1938). Skinner could not have taught such a complex sequence of responses all at once. Instead, he added “links” to the chain of learned responses one at a time: he first trained the rat to pull the string, then trained it to pull the string and pick up the marble, then trained it to pull the string and pick up the marble and carry it to the tube, and so on. Sometimes, it is more effective to train the steps in reverse order, in a process called backward chaining: first train the rat to drop the marble in the tube, then train the rat to carry the marble to the tube and drop it in, and so on. Either way, at each stage, the rat must perform a progressively longer sequence of responses to gain its food.

Chaining is a useful technique for training humans, too. Workers learning a manufacturing process are often taught it one step at a time (Walls, Zane, & Ellis, 1981), and trainee pilots may master landing sequences by practicing progressively longer sequences on a flight simulator (Wightman & Sistrunk, 1987).

Learning About Outcomes

So far we have discussed outcomes fairly loosely. Formally, a reinforcer is a consequence of behavior that leads to increased likelihood of that behavior in the future. For example, food is a reinforcer to a hungry animal, and the animal may learn to repeat behaviors that result in access to food. On the other hand, a punisher is a consequence of behavior that—you guessed it—leads to decreased likelihood of that behavior in the future.
LEARNING AND MEMORY IN EVERYDAY LIFE

Drug-Detecting Dogs

If you’ve traveled by air in the past few years, then your luggage has probably been inspected by specially trained sniffer dogs, now commonly used (often behind the scenes) at airports and border crossings to help authorities detect smuggled drugs, explosives, and other illegal substances. The sniffer dogs can search baggage in a fraction of the time it would take a human. And the dogs are not only fast but sensitive; they can detect faint odors even if the target substance is sealed inside plastic or smeared with chocolate to disguise the scent.

Sniffer dogs are taught their jobs through operant conditioning. By now, you should be able to guess many of the details. Training usually starts with positive reinforcement, in which a puppy learns to fetch a toy to obtain a reward such as playtime with the trainer (a reward most dogs find highly reinforcing). Then the toy is doused with a target scent—say, the scent of heroin. Each time the dog retrieves the scented toy and gets a rewarding bout of playtime, the association between odor, retrieval, and reward is strengthened.

The trainers continue by introducing other objects doused with the same odor, so that the dog learns to retrieve any object with the target scent. Sometimes a clicker or other secondary reinforcer (defined on page 185) is introduced, so that the training session need not be interrupted for playtime after every successful retrieval. Later on, the dogs may be trained to give an “alert” response instead of actually retrieving the object, and the target scent can be expanded to include multiple odors, so that a dog will give the alert response upon sensing cocaine and other illicit drugs in addition to heroin.

How accurate are the dogs? One study of more than 1,200 tests performed in police canine training sessions reported that, on average, the dogs correctly indicated the presence of drugs in more than 90% of trials that involved detecting marijuana but only about 70% of trials with heroin (Jezierski et al., 2014). The dogs also produced false alarms in about 12% of heroin trials, giving the alert response when no drugs were present.

The false alarm rate may be even higher in “real-world” searches. A widely publicized report by the Chicago Tribune claimed that when police used drug-sniffing dogs to check cars pulled over during traffic stops, the dogs generated more false alarms than correct responses (Hinkel & Mahr, 2011). Such false alarms can have serious economic consequences, as when an airport is shut down for several hours after a bomb-detecting dog identifies a suspicious object that is later determined to be harmless.

Why so many false alarms? Part of the reason may simply be that drug odors can persist for up to 48 hours (Jezierski et al., 2014), so the dogs may be responding to lingering scents from drugs that are no longer present. But another possibility is that dogs respond to subtle (possibly unconscious) cues from their human handlers. In one laboratory study, where the handlers had been led to believe that the target odor was hidden in a location marked by a piece of red paper, the dogs gave significantly more alert responses in that location, even though no odor was present (Lit, Schweitzer, & Oberbauer, 2011). Furthermore, when a dog gives an alert signal in the field, the attention this signal elicits from the handler may unintentionally reinforce the signaling behavior, even if the alert later turns out to have been a false alarm.

Interestingly, similar controlled studies with giant African pouched rats (such as the mine detectors you read about in the chapter opener) suggest that the rats are much less prone to false alarms. A possible explanation is that, while dogs are highly motivated to interact with and obtain attention from their human handlers, the pouched rats seem less interested in earning praise from their handlers and more focused on how to earn the next banana reward.

**reinforcement** The process of providing outcomes (reinforcers) that lead to increased probability of a particular behavior occurring in the future.

**punishment** The process of providing outcomes (punishers) that lead to decreased probability of a particular behavior occurring in the future.

For instance, a rat will learn to stop pressing a lever if it receives a mild electric shock for doing so. Some children who have not stopped thumb sucking on their own agree to have their thumbs coated with a harmless but bitter-tasting substance; each thumb-sucking response then results in a nasty taste, and this punishment may decrease the rate of thumb sucking in the future.

Put another way, **reinforcement** is the process of providing outcomes (reinforcers) that lead to increased probability of a behavior, and **punishment** is the process of providing outcomes (punishers) that lead to decreased probability of a behavior. Sounds simple enough, right? But the reality is more complicated, for both reinforcers and punishers.
Primary Reinforcers

Food, water, sleep, comfortable temperatures, and sex are all examples of primary reinforcers: they are of biological value to organisms, and therefore organisms will tend to repeat behaviors that provide access to these things or conditions. Psychologist Clark Hull’s drive reduction theory proposed that all learning reflects the innate, biological need to obtain primary reinforcers (Hull, 1943, 1952). The motivation to obtain primary reinforcers was a key variable in Hull’s equations—with which, as discussed in Chapter 1, he hoped to explain all learning.

One complication to this theory is that primary reinforcers are not always equally reinforcing. Thirsty animals will work to obtain access to water, a primary reinforcer; but if they are not thirsty, access to water is not particularly reinforcing. In addition, not all primary reinforcers are created equal. Hungry animals will work to obtain food, but they will work even harder for food they like. For example, rats will run a maze faster for bread and milk (which they find especially tasty) than for sunflower seeds, even though the seeds satiate hunger just as effectively (Simmons, 1924).

In short, the identity of the reinforcer matters. In operant conditioning, organisms learn that the response R results not just in any random outcome but in a specific outcome O. A switch in the outcome may produce a change in responding. For instance, rats can be trained to make lever-press responses to obtain either food pellets or sugared water, but they tend to prefer the sugared water. If rats are first trained to respond in order to obtain sugared water, and then the reinforcer is switched to food pellets, the rats’ rate of responding plummeted (Weatherly, Plumm, Smith, & Roberts, 2002). This is an example of negative contrast, in which the reinforcing value of one reward is reduced because a better reward is expected (Flaherty, 1982).

Negative contrast can be observed in humans, too. In one classic study, infants were given access to an artificial nipple that they could suck to obtain water. Some infants received plain water in sessions 1 and 2 (Figure 5.3, light blue line), while others received sugared water in both sessions (dark blue line). Unsurprisingly, the infants who were given sweetened water sucked more vigorously: plain water is an acceptable reinforcer, but sweetened water is a preferred reinforcer. But a third group of infants was given sweetened water in session 1 and then switched to plain water in session 2. As shown in Figure 5.3 (orange line), these infants’ responses plummeted after the switch—to a lower level than those infants who had received plain water all along (Kobre & Lipsitt, 1972). Similar negative contrast effects can be observed in children who go trick-or-treating expecting candy but instead get raisins (which on normal days they might enjoy), as well as in game-show contestants who hope to win the million-dollar grand prize but instead dejectedly settle for the consolation prize of a free vacation, which they would normally be delighted to accept.

Secondary Reinforcers

In addition to primary reinforcers such as food and water, learning can also be driven by secondary reinforcers, which are reinforcers that initially had no biological value but have been paired with (or predict the arrival of) primary reinforcers (Shahan, 2010). The best example of a secondary reinforcer is money. Money itself has no biologically reinforcing properties, but it can be exchanged for any number of primary reinforcers, including
Secondary reinforcement is often used in prisons, psychiatric hospitals, and other institutions to motivate inmates or patients to respect rules or perform chores such as making beds. Each desired behavior is reinforced with a token, and tokens can then be exchanged for privileges (say, access to the telephone or group activities). Such arrangements are called token economies because the tokens function the same way as money does in the outside world (Hackenberg, 2009). Token economies have been used with some success to modify behavior in children with intellectual disability or with autism (Matson & Boisjoli, 2009); even nonverbal children may be able to learn that certain responses result in acquisition of tokens, which in turn can be exchanged for candy or toys.

Animals, too, will work for secondary reinforcers. For example, trainers can use secondary reinforcement to teach dolphins to do tricks. The trainer first pairs a whistle sound with food reinforcement until the dolphin has learned an association between the whistle and food; at this point, the whistle has become a secondary reinforcer, and it can be used to maintain the behavior (Pryor, Haag, & O'Reilly, 1969). Horse trainers often use a similar technique, pairing a clicking noise with oats and then eventually using the clicking noise alone (Skinner, 1951). In both cases, as long as the secondary reinforcer is occasionally followed by food, the behavior is maintained. Secondary reinforcers are particularly useful in animal training because the animal trainer can deliver clicker reinforcement immediately and from a distance; the more quickly the response is followed by reinforcement, the more effective the reinforcement is.

A standard explanation of secondary reinforcement is that, by virtue of being paired with a primary reinforcer (such as food), secondary reinforcers (such as the clicker) become reinforcers themselves, which the organism will work to obtain. Some evidence supports this view. However, other research suggests that animals are not “fooled” into thinking that the secondary reinforcers are worth working to obtain; rather, the secondary reinforcers provide informational feedback that behavior is on the right track for obtaining the primary reinforcer: “Keep executing this response, and you’ll eventually get food” (Shahan, 2010). This is similar to a prison inmate who completes chores not because she is particularly motivated to obtain the tokens themselves but because she knows that acquiring tokens is a way to obtain things that she does like.

Remember the giant African pouched rats, mentioned in the introduction to this chapter, that have been trained to sniff out buried mines? They are trained through a combination of secondary reinforcement and shaping (Poling et al., 2010, 2011). Young rats are trained (through classical conditioning) to associate the popping sound of a clicker with a small reward (such as banana pulp). The food is a primary reinforcer, and the clicker sound becomes a secondary reinforcer that predicts the food. Next, the rats are placed in a cage that contains a hole in the floor, below which the scent of an explosive (such as TNT) is placed. The rats are trained to hold their noses in the hole for a few seconds in order to earn clicker reinforcement. Then the rats are trained with multiple holes, only some of which are TNT-scented, following which the rats are
reinforced for correctly performing the “sniff and hold” response to the scented holes but not to the unscented ones:

- $S^D$ (odor of TNT) $→$ R (sniff and hold) $→$ O (clicker)
- $S^D$ (no odor of TNT) $→$ R (sniff and hold) $→$ O (no clicker)

The clicker is particularly useful here because the trainers can provide the reinforcement immediately, without waiting for the rat to finish responding and retrieve a treat. The clicker sound itself is still occasionally paired with food so that it remains reinforcing.

Eventually, the rats are moved outside, where they can run around small yards that contain buried TNT-scented targets, and finally they are trained to work methodically up and down marked paths in an actual simulated minefield. Only after a rat can successfully detect every simulated mine in a 200-square-meter test area containing 5 to 7 targets is the animal “licensed” as an operational mine-detection animal.

The same general techniques are used to train giant pouched rats to detect tuberculosis in human sputum samples, as described at the start of this chapter. The key here is that infected samples emit an odor that the rats’ highly sensitive olfactory system can recognize. The odor of infection becomes the discriminative stimulus $S^D$, triggering the “sniff and hold” response R, and resulting in clicker reinforcement O (which is sometimes paired with food):

- $S^D$ (odor of infected sample) $→$ R (sniff and hold) $→$ O (clicker)

After several weeks of training, rats can move rapidly down a row of sample wells, pausing for clicker reinforcement when they sniff out an infected sample—that is, signaling to the human observer which of the samples likely came from infected patients. Whereas human microscope technicians can process only a few dozen samples per day, highly trained rats can process (sniff at) 140 samples in about 20 minutes, and studies suggest that the rats’ diagnoses are more than 90% correct (Weetjens et al., 2009). As mentioned at the beginning of the chapter, trained rats are currently being used in clinics in Tanzania and Mozambique, and some studies suggest they can even identify infected samples that are “missed” by standard lab tests alone (Mgode et al., 2018).

**Punishers**

Just as organisms will work to obtain reinforcers, they will also work to avoid punishers. Common punishers for animals include pain, confinement, and exposure to predators (or even the scent of predators). Common punishers for humans are monetary fines, social disapproval, and jail time.

As discussed in Chapter 1, Thorndike (1911) originally assumed that punishers were simply the inverse of reinforcers: whereas reinforcement increases the probability that a response will occur again in the future, punishment decreases that probability. Later, both Thorndike (1932) and Skinner (1938, 1953) concluded that punishment is not nearly as effective as reinforcement at controlling behavior. In the end, Thorndike dropped the idea of punishment from his law of effect, concluding that while reinforcement does indeed increase the probability that a response will be repeated in the future, the effects of punishment are erratic and unreliable and
can even at times result in paradoxical increases in the punished behavior (Postman, 1962; Thorndike, 1943).

Nevertheless, many modern researchers argue that punishment can indeed be very effective in modifying behavior (for review, see Staddon, 1995). The problem is that several factors determine how effective the punishment will be. We describe four of the most important factors here.

1. **Punishment leads to more variable behavior.** According to the law of effect, reinforcement of a particular response R increases the probability that the same response R will occur in the future. In contrast, punishment of R decreases the probability that R will occur in the future but does not predict what response will occur instead of R! In fact, punishment tends to produce variation in behavior, as the organism explores other possible responses. That is fine if the primary goal is simply to eliminate an undesired response (such as training a child not to go near a hot stove). But it is not a particularly good way to train desired behaviors. If the goal of conditioning is to shape behavior in a predetermined way, then reinforcing the desired response generally produces much faster learning than simply punishing alternative, undesired responses.

2. **Discriminative stimuli for punishment can encourage cheating.** Discriminative stimuli can signal whether an operant response will be reinforced, and they can also signal whether a response will be punished. This means that the absence of the discriminative stimulus might signal that the response will not be punished. For example, to a speeding driver, the sight of a police car is a discriminative stimulus for punishment: speeding in the presence of this stimulus will probably be punished. But speeding in the absence of a police car will probably not be punished. In this case, punishment does not train the driver to drive slowly; it only teaches him to suppress speeding in the presence of police cars. When no police car is visible, speeding may resume. Similarly, the dominant male in a group of chimpanzees may punish females for mating with any other males—but when his back is turned, females often sneak off into the bushes with lower-ranking males. And rats that have been painstakingly trained to eat no more than four pellets at a time will often eat all the food in sight if no human is watching (Davis, 1989).

3. **Concurrent reinforcement can undermine the punishment.** The effects of punishment can be counteracted if reinforcement occurs along with the punishment. Suppose a rat first learns to press a lever for food but later learns that lever presses are punished by shock. Unless there is another way to obtain food, the rat is likely to keep pressing the lever to obtain food reinforcement, in spite of the punishing effects of shock. Similarly, a child may be reprimanded for talking in class, but this punishment will be much less effective if the behavior is simultaneously reinforced by approval from classmates. And although a speeding driver risks a hefty ticket, the effects of this punisher may be countered by the reinforcing fun of driving fast.

4. **Initial intensity matters.** Punishment is most effective if a strong punisher is used from the outset. In one study, rats received a shock as they ran through a maze to the goal box (Brown, 1969). The shock was initially delivered at the lowest intensities (1 or 2 volts) and had little effect on behavior. Gradually, across several days, shock intensity increased to 40 volts. Behavior was essentially unaffected in these rats, even though naive rats given a 40-volt shock would stop running immediately.
Apparently, early weak shocks made the rats insensitive to later, stronger shocks. The effectiveness of the strong shock was completely undermined by starting weak and working up from there. Similarly, a child who misbehaves in class may first receive a warning, then a scolding, then detention, then expulsion. The expulsion, when it comes, may be much less effective at deterring future misbehavior because of the prior milder punishments. Likewise, a speeding driver may not be deterred much by a hefty $500 ticket if she is already accustomed to paying lesser fines. In each case, the prior weak punishers may undermine the effectiveness of the severe punisher, when it finally comes. (See “Learning and Memory in Everyday Life” below for more discussion of the problems with punishment.)

**Differential Reinforcement of Alternative Behaviors**

Given the problems with punishment, many prefer using the carrot rather than the stick. Instead of delivering punishment each time an unwanted behavior is exhibited, one can reward preferred alternative behaviors—a process known as **differential reinforcement of alternative behaviors** (abbreviated DRA). For example, some children with autism or developmental disorders show persistent habits of self-injurious behavior, such as repeatedly banging their head against the wall or biting their own hand. Rather than punishing the child for each instance of the unwanted behavior, parents or therapists can reward instances of desired behavior, such as compliance.

---

**LEARNING AND MEMORY IN EVERYDAY LIFE**

**The Problem with Punishment**

Around the world, about 80% of children are occasionally spanked or otherwise physically punished by their parents as a way to discourage unwanted behaviors (UNICEF, 2014). Spanking is a form of punishment that even a very young child can understand; it is easy to administer, and it can effectively reduce the frequency of an unwanted response.

But physical punishment is controversial. Many people believe that hitting a child is never justifiable. Some studies have suggested that children who are spanked can develop emotional problems, including aggression and stress (Gershoff & Grogan-Kaylor, 2016; Gershoff, Sattler, & Anasari, 2017), whereas other studies have found that occasional mild spanking does not cause any lasting harm (Ferguson, 2013).

Parents who want to avoid spanking have other options. Punishment does not have to cause physical pain to be effective. Scolding is a form of punishment that does not cause physical harm; other methods are time-out, grounding, and withholding of allowance.

But there is still the problem that punishment is hard to apply effectively. Take a hypothetical example. Shawn has two busy working parents and several older siblings. When Shawn is well behaved, his siblings tend to get most of the parental attention; but when Shawn breaks china, fights with his brothers, or causes trouble at school, the parental spotlight shines on him. Although they may think they are punishing Shawn, his parents are actually reinforcing his bad behavior by giving him attention when he misbehaves.

What can be done? For one thing, Shawn’s parents should punish unwanted behavior with a minimum of fuss so that the offender gets less attention—and less reinforcement—for it. They can minimize opportunities for the unwanted behavior (such as by keeping the china on a higher shelf). They can also reduce unwanted behavior by differential reinforcement of alternative behaviors, praising Shawn on the days when he does not misbehave. This means that Shawn’s parents have to commit the extra time and effort to pay more attention to their youngest child, not by merely punishing him when he is bad but by also acknowledging him when he is good. The payoff may be a well-behaved child and a happier family, while avoiding many of the problems of punishment.
with instructions to complete a task or eat a nourishing meal (Petscher, Rey, & Bailey, 2009). DRA can work particularly well if the rewarded behavior is incompatible with the unwanted behavior. For instance, as long as she is sitting in a chair to complete homework or eat a meal, the child is physically unable to bang her head against the wall.

In 2011, the movie theater chain Cinemark introduced a mobile phone application that allows customers to check movie listings and buy tickets. It also tried a new tactic to combat a pernicious feature of the modern cinema experience: fellow moviegoers who text during the performance, creating phone screen light that may distract those seated nearby. Warnings at the start of the film to put away cell phones, threats to eject the disobedient, and even social disapproval all having failed, Cinemark turned to reinforcement of an alternate behavior: users who activate the phone app at the start of the movie and don’t use the phone again for the rest of the movie earn coupons—secondary reinforcers that can be exchanged for rewards like free popcorn and soda. The jury is still out on whether this attempt at DRA actually reduced texting-while-viewing, compared with more traditional punishment-based approaches.

### Putting It All Together: Building the S^D → R → O Association

The components of operant learning—the discriminative stimulus, response, and outcome—can be arranged in different ways. An experimenter can vary factors such as the temporal spacing between R and O, whether the outcome is something added to or subtracted from the environment, and even the regularity with which the outcome follows the response.

### Timing Affects Learning

In most of the operant conditioning examples presented so far, the outcome (reinforcer or punisher) has immediately followed the response. For instance, as soon as the rat presses a lever, food drops into the Skinner box; as soon as a pouched rat locates a buried land mine, the trainer provides clicker reinforcement; as soon as toddler Annie starts to suck her thumb, the bitter taste on her finger reminds her to stop.

Normally, immediate outcomes produce the fastest learning. If there is no delay between response and reinforcement, then the odds are good that the most recent behavior will be identified as the response that caused the outcome, and the frequency of that response will increase. But if there is a long delay, it is more likely that other behaviors have taken place during the interval, and those behaviors (instead of the earlier response) might thus be associated with the outcome. To illustrate this idea, Figure 5.4
shows that rats learn a lever-pressing task quickly when the delay between response and food delivery is 0 seconds, but they are slower to learn the association if the delay is 4 seconds, and even slower if the delay is lengthened to 10 seconds (Schlinger & Blakely, 1994).

Temporal contiguity of response and outcome has a similar impact on the effectiveness of punishment. Unfortunately, human society often employs delayed punishment. Criminals may not come to trial—much less serve their sentence—until months or years after committing the crime. A middle school student who misbehaves in the morning and receives detention after school experiences a delay of several hours between response and outcome. Such delays undermine the punishment’s effectiveness and may weaken learning.

**Outcomes Can Be Added or Subtracted**

Suppose Becky learns that, at dinnertime ($S^D$), helping to set the table (R) will result in the reinforcement of parental praise (O). Her younger sister Annie learns that, after a bitter coating has been applied to her thumb ($S^D$), sucking her thumb (R) will trigger an unpleasant taste (O). Note that in both these examples, the outcome is something “added” to the child’s environment:

- $S^D$ (dinnertime) → R (set the table) → O (praise)
- $S^D$ (tastant on thumb) → R (thumb sucking) → O (bitter taste)

These paradigms are technically called **positive reinforcement** and **positive punishment**. Note that here the word *positive* does not mean “good”; instead, it means “added” in the mathematical sense (as in a positive number). In positive reinforcement, the desired response causes the reinforcer to be added to the environment; in positive punishment, an undesired response causes a punisher to be added to the environment.

There are also learning situations in which the outcome is the removal of something from the environment. In **negative reinforcement**, behavior is encouraged (reinforced) because it causes an undesirable element to be removed from the environment. Here, the word *negative* means “subtracted” in the mathematical sense (as in a negative number). For example, if you have a headache, you can take aspirin to make the headache go away:

- $S^D$ (headache) → R (take aspirin) → O (no more headache)

The net result is that you are more likely to take aspirin again the next time you have a headache, so this scenario is an example of reinforcement. The outcome is not something added to your environment but something subtracted—that is, the headache is taken away—so this is negative reinforcement.

Similarly, a rat can be placed in a chamber with an electrified floor grid that administers electric shocks. The rat can escape these shocks by climbing onto a wooden platform. In this case, the response is climbing, and the outcome is escaping from the shock:

- $S^D$ (shock) → R (climb) → O (no more shock)

The net result is that the rat is more likely to climb the platform in the future. In other words, the climbing response has been reinforced. Because the outcome involves a subtraction (i.e., shock is subtracted, or taken away), this is an example of negative

---

**positive reinforcement**
A type of operant conditioning in which the response causes a reinforcer to be “added” to the environment; over time, the response becomes more frequent.

**positive punishment**
A type of operant conditioning in which the response causes an undesirable element to be “added” to the environment; over time, the response becomes less frequent.

**negative reinforcement**
A type of operant conditioning in which the response causes an undesirable element to be “subtracted from” the environment; over time, the response becomes more frequent.
negative punishment. A type of operant conditioning in which the response causes a desirable element to be “subtracted from” the environment; over time, the response becomes less frequent.

reinforcement. Negative reinforcement is sometimes called escape or avoidance training because the response causes an escape from, or avoidance of, something aversive (such as a headache or shock).

Just as behavior can be reinforced by taking away bad things, so can behavior be punished by taking away good things. This kind of paradigm is called negative punishment because something is subtracted (taken away) from the environment, and this subtraction punishes the behavior. (Negative punishment is also sometimes called omission training because the response R results in something being “omitted from” the environment.) For example, if Becky displays aggressive behavior toward other children during recess, the teacher may make Becky sit by herself while the other children play:

\[ S^D (\text{recess}) \rightarrow R (\text{aggressive behavior}) \rightarrow O (\text{loss of playtime}) \]

The net effect is that Becky may be less likely to display aggressive behavior in the future. This kind of negative punishment is sometimes called a time-out: Becky is punished by time away from a normally reinforcing activity. Time-outs work only if the activity being restricted is something reinforcing. A time-out from an activity the child dislikes may actually serve to reinforce, rather than reduce, the bad behavior that earned the time-out!

Negative punishment is widely applied in human society: teenagers may be grounded for staying out too late, drivers may have their licenses suspended for speeding, and customers who do not pay their phone bills on time may have their service discontinued. In each case, an undesirable behavior is punished by revoking privileges, in the hope of decreasing the likelihood that such behavior will occur again in the future.

Table 5.1 summarizes the four types of training. Keep in mind that the terms reinforcement and punishment describe whether the response increases in frequency (reinforcement) or decreases in frequency (punishment) as a result of training. The terms positive and negative describe whether the outcome is something added (positive) or something taken away (negative).

Laboratory experiments often fit neatly into the grid shown in Table 5.1, but real life is more complicated. Sometimes it is difficult to determine whether an individual is learning based on reinforcement or punishment or both. For example, when studying for an exam, are students working to obtain a good grade (positive reinforcement) or to avoid flunking (negative reinforcement)? It could be either—or both. Similarly, a

<table>
<thead>
<tr>
<th>Outcome is added (positive)</th>
<th>Response increases (reinforcement)</th>
<th>Response decreases (punishment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive reinforcement</td>
<td>Example: Clean room → get weekly allowance</td>
<td>Positive punishment Example: Exceed the speed limit → ticket and fine</td>
</tr>
</tbody>
</table>

| Outcome is removed (negative) | Negative reinforcement (escape/avoidance training) Example: Take aspirin → headache goes away | Negative punishment (omission training) Example: Fight with other children → time-out from play |
child who misbehaves may receive both a scolding (positive punishment) and a time-out (negative punishment), and both considerations may motivate her to keep bad behavior to a minimum.

## TEST YOUR KNOWLEDGE

### Reinforcement Versus Punishment

It is easy to confuse the ideas of negative reinforcement, positive punishment, and so on because in other contexts we often use the words *positive* and *negative* to mean “good” and “bad.” Don’t fall into this trap! First ask yourself whether the outcome is added to (positive) or subtracted from (negative) the environment, and then ask yourself whether the outcome causes a behavior to increase (reinforcement) or decrease (punishment).

Try your hand at each of the following scenarios and see if you can tell whether it is an example of positive reinforcement, negative reinforcement, positive punishment, or negative punishment. In each case, answer these questions in order to decide: (a) Who does the learning? (b) What is the response (the behavior that is being altered)? (c) What is the outcome? (d) What is the discriminative stimulus that determines whether the response will produce that outcome? (e) Is the outcome something added or taken away? (f) Does the response increase or decrease as a result of learning? *(Answers appear at the end of the chapter.)*

1. At the grocery store, 2-year-old Lucy sees candy and wants it. Her mother Susan says no, and Lucy throws a temper tantrum. To calm her down, Susan relents and buys Lucy some candy. The next time they go shopping, Lucy sees candy and immediately throws another tantrum. This time, she obtains the candy quickly.

2. An interesting aspect of operant conditioning is that sometimes more than one actor is doing the learning. Scenario 1 is presented from Lucy’s point of view. But consider the same story from Susan’s point of view: Susan takes her toddler on a shopping trip. The child sees candy, wants it, and throws a tantrum. Overtired and in a rush, Susan gives the child some candy, and the tantrum stops. On the next trip, as soon as the child starts a preliminary wail, Susan quickly hands over some candy to stop the screaming.

3. Shevonne installs an electric fence system around the perimeter of her yard and gives her dog Snoopy a collar that makes a high-pitched noise whenever he gets too close to the boundary. The first time Snoopy strays out of bounds while wearing the collar, the noise plays and distresses him. Soon, Snoopy learns to avoid the noise by staying inside the yard.

4. Miguel’s football team has a no-alcohol policy: players sign pledges not to drink alcohol during the football season. One night, Miguel goes out with some friends and has a few beers. The coach finds out and revokes Miguel’s playing privileges for a week. When allowed to rejoin the team, Miguel is careful to stay away from alcohol for the rest of the season.

5. Rachel is a 10-year-old who hates gym class. One day, after eating the school lunch, she gets a stomachache. She tells the school nurse she is feeling sick, and the nurse gives her a pass to skip gym class that afternoon. Now, on days when there is gym class, Rachel frequently feels sick after eating lunch at school.

### Reinforcement Need Not Follow Every Response

In addition to controlling whether a response results in reinforcement or punishment being added or taken away, and whether the outcome ensues immediately or after a delay, an experimenter can also control the *frequency* with which these outcomes are delivered.
The rules determining how and when outcomes are delivered in an experiment are called reinforcement schedules. So far, almost all of the examples in this chapter have been ones in which the outcome reliably follows the response. For instance, whenever the rat presses a lever, it gets food; whenever Becky sets the dinner table, she receives parental praise; and so on. These examples illustrate continuous reinforcement schedules, meaning that each response R is always followed by the outcome O.

But in some situations, a response must be repeated multiple times before triggering the expected outcome. For example, Becky has to clean her room seven days in a row to obtain her weekly allowance (seven responses for one reinforcement), and a baseball player is allowed to swing and miss three times before he strikes out (three responses for one punishment).

In the laboratory, an experimenter can devise a schedule defining exactly when outcomes are delivered. Patterns in which an outcome follows a response less than 100% of the time are called partial reinforcement schedules (or intermittent reinforcement schedules). The term reinforcement schedule is used here for simplicity, but these schedules can be applied either to reinforcement (Becky’s weekly allowance) or to punishment (three strikes and you’re out).

Reinforcement Schedules

There are four commonly used types of partial reinforcement schedules, defined by the specific rules determining when reinforcement arrives.

1. Fixed-ratio (FR) schedule. In an FR schedule, some fixed number of responses must be made before a reinforcer is delivered. For example, if a rat must press a lever five times to obtain one food pellet, the ratio of responses to reinforcers is 5:1; this is often called an FR 5 schedule. In this notation system, continuous reinforcement would be considered an FR 1 schedule: every (one) response results in reinforcement. The ratio can gradually be increased—for instance, by starting with an FR 1 schedule and working up through an FR 5 schedule to an FR 50 schedule and beyond—until the animal has been trained to make several hundred responses for each reinforcement on an FR schedule.

Rats on an FR schedule show a characteristic pattern of steady responding leading up to the reinforcement, followed by a brief pause with no responding. This short break in responding after a reinforcement is called the post-reinforcement pause. This pattern of responding under an FR schedule is most easily seen using a graph of cumulative responding (the type introduced in Figure 5.2D), which is one reason some researchers continue to display data in this format. Figure 5.5A (blue line) shows the hypothetical behavior of a rat trained to respond on an FR 5 schedule: steady response rates leading up to each reinforcement (orange arrow), followed by a brief pause before another round of responding begins in order to obtain a new reinforcement.

During the post-reinforcement pause, it seems almost as if the animal is pausing to take a rest before its next bout of responding. And, in fact, the length of the post-reinforcement
pause is related to the number of responses required to obtain the next reinforcement: thus, the post-reinforcement pause when an animal is on an FR 50 schedule is longer than when the animal is on an FR 5 schedule. In effect, the rat is behaving like a human teenager who completes a quick chore (say, taking out the trash) the first time his mother asks but procrastinates for hours before starting a really time-consuming chore (mowing the lawn).

Examples of fixed-ratio schedules in human life include factory workers getting paid a flat fee for every 100 pieces they turn out and farm workers getting paid a fixed amount for every bushel of apples picked. In fact, such workers tend to show behavior similar to that of rats on an FR schedule—steady bursts of responding followed by post-reinforcement pauses: the workers complete a batch, take a few minutes for a coffee break, and then start in again. A similar phenomenon occurs in readers: they may complete a chapter or a fixed number of pages before putting the book aside. Novelists often try to combat this “post-reinforcement pause” by ending each chapter with an exciting cliffhanger so that readers will read on to see what happens next.

2. **Fixed-interval (FI) schedule.** Whereas an FR schedule provides reinforcement after a fixed number of responses, an FI schedule reinforces the first response after a fixed amount of time. For example, on an FI 10-sec schedule, the rat is reinforced for the first response it makes after a 10-second interval since the last reinforcement. Importantly, the reinforcement is not automatically obtained after the fixed interval but merely becomes available to be earned; the organism must still respond in order to actually receive that reinforcement. Once the interval has elapsed, the reinforcement remains available until the response occurs and the reinforcement is obtained. At that point, the clock starts ticking on the next fixed interval until the next reinforcement becomes available.

Under these circumstances, the most efficient strategy on the rat’s part would be to wait exactly 10 seconds after each reinforcement and then respond once to obtain the next reinforcement. Earlier responses (before the 10 seconds have elapsed) are wasted...
effort. However, Figure 5.5B shows how an animal on an FI schedule actually behaves: each reinforcement is followed by a period of few or no responses, but the animal’s rate of responding gradually increases as the end of the interval nears. Presumably, animals (including humans) cannot perfectly judge time intervals, so they estimate as best they can how much time has passed and err a little on the side of wishful thinking by responding too soon. The result is a characteristic “scalloped” cumulative response curve, as illustrated in Figure 5.5B (blue line).

An example of fixed-interval reinforcement would be a high school student being sentenced to serve detention from 3 p.m. to 4 p.m. After he arrives, there is little point in checking the clock for the first 15 or 20 minutes. But it might be worth checking after he estimates 30 to 40 minutes have passed, just in case time is going by faster than he thinks. As the elapsed time gets closer and closer to an hour, he might check the time more and more frequently, not wanting to stay a moment longer than necessary. In this example, the response is checking the time, the reinforcement is escape from detention, and only the first response after the end of the detention interval is reinforced—the rest are “wasted” responses. Notice that this is an FI schedule, not an FR schedule, because the rate at which the student checks his watch does not make the reinforcement appear any faster. Rather, the reinforcement arrives following the first response after the time interval has elapsed. Once 4 p.m. arrives, however, the student does not obtain reinforcement until he checks the clock and realizes he is free to go.

3. **Variable-ratio (VR) schedule.** A VR schedule provides reinforcement after a certain average number of responses. For example, whereas an FR 5 schedule produces reinforcement after every fifth response, a VR 5 schedule produces reinforcement after every five responses on average. Thus, the responder never knows exactly when a reinforcement is coming. As a result, there is a steady, high rate of responding even immediately after a reinforcement is delivered because the very next response just might result in another reinforcement (Figure 5.5A, green line). Thus, the VR schedule eliminates (or greatly reduces) the post-reinforcement pause observed under FR schedules.

A real-life example of a VR schedule could be a slot machine. Even if you know that a particular slot machine pays off on every 500th game, on average, you do not know exactly which games will pay off. Even if you have just hit the jackpot, the very next spin might be a winner, too, so there is a strong incentive to keep playing.

4. **Variable-interval (VI) schedule.** Whereas an FI schedule reinforces the first response after a particular time interval, a VI schedule reinforces the first response after an interval that averages a particular length of time. For example, a VI 10-sec schedule reinforces the first response after an interval that is 10 seconds on average—but the actual interval on any particular trial might be longer or shorter.

In the VI schedule, as in the VR schedule, the responder never knows exactly when the next reinforcement is coming: responses made a few seconds after the previous reinforcement just might be reinforced, too. Thus, the response rate of animals under a VI schedule is usually steadier than under an FI schedule, as the animals check periodically to see whether reinforcement is available (Figure 5.5B, green line).
An everyday example of the effect of a VI schedule is demonstrated by Tom, a college student whose girlfriend is studying abroad for a semester. She has promised to keep in touch through daily online posts, but sometimes she posts in the morning, sometimes in the evening—and sometimes more than once a day. Tom loves her and wants to read her posts as soon as they arrive, but he also does not want to spend 24 hours a day staring at his screen and waiting for the next post to come. A sensible compromise is for Tom to check online every few hours. This minimizes the time unread posts will sit waiting for him, while leaving him free to go pursue other interests in the meantime. Notice that this is a VI schedule, not a VR schedule, because the rate at which Tom checks online does not make the posts appear any faster. Once a new post does appear, it sits there waiting until Tom checks; his next response is rewarded, as he gets to read the post.

Just as VR schedules tend to produce higher rates of responding than FR schedules, so VI schedules tend to produce higher rates of responding than FI schedules. For example, if Tom’s girlfriend posts every day at the fixed time of 2 p.m., Tom will tend to check online once every day at about that time; if she posts at variable times, he is likely to check much more often, just in case a new post is waiting. One reason many people check their phones every few minutes is that texts and tweets arrive on a VI schedule: you can never be sure when there is a juicy new item sitting there, waiting to reinforce your behavior.

**TEST YOUR KNOWLEDGE**

**Reinforcement Schedules**

Operant conditioning is common in human behavior. Try your hand at identifying whether each of the scenarios below is an example of an FR, FI, VR, or VI schedule. *(Answers appear at the end of the chapter)*

1. Each first-grade student who completes the day’s math worksheet receives a gold star from the teacher; at the end of the week, five gold stars can be exchanged for a toy.
2. A good telemarketer scores an average of two sales for every 20 phone calls he makes, so he earns the most profit if he makes a lot of calls.
3. A couple go to their favorite restaurant on a Saturday night and are told that seating will be available in about 30 minutes. They wait at the bar and periodically return to the reception area to check whether a table is free.
4. Maria donates blood regularly at the local hospital; they pay her for her donations, and it makes her feel good to know she’s helping people in need. However, due to hospital policy, donors have to wait at least 2 weeks between donations.
5. A surfer spends all available afternoons at his favorite beach, where he is sure of at least a couple of big waves every hour or so. After catching a big wave, he immediately paddles back out to await the next big one.
6. A man who likes to eat spicy foods for lunch always carries a pack of spearmint chewing gum so he can freshen his breath before returning to work in the afternoon.
7. A calculus teacher gives quizzes every Friday. Some students start looking over the material early in the week, but everyone in the class studies frantically on Thursday night. Usually, those who spend the most time studying get the best grades on the quiz.
8. A woman likes to play bingo at her local church. The game is set up so that 1 out of every 100 cards will be a winner, but it is impossible to know in advance which specific cards will win. To increase her chances of winning, the woman buys 10 cards each time she plays.
Choice Behavior

In addition to continuous and partial reinforcement schedules, there are also concurrent reinforcement schedules, in which the organism can choose between several possible responses, each leading to a different outcome. This allows researchers to examine choice behavior: how organisms choose to divide their time and efforts among different options. For example, suppose a pigeon is placed in a Skinner box with two illuminated disks, A and B (Figure 5.6A). Pecking at disk A is reinforced on a VI 2-min schedule (2-min = 2-minute), and pecking at disk B is reinforced on a VI 1-min schedule. What should the pigeon do?

Every 2 minutes, the pigeon can get one food pellet for pecking at disk A but gets two for pecking at disk B—so you might think that the pigeon would concentrate on pecking at B and ignore A. But remember, if the experiment lasts longer than 2 minutes, a food pellet will be waiting to be delivered as soon as the pigeon pecks at A, and this pellet will never be obtained if the animal pecks only at B. The optimal behavior is some strategy that allows the pigeon to maximize the amount of food it can get from both disks, probably by spending the most effort pecking at B but occasionally switching over to peck at A, just to check. This would be analogous to channel surfing in humans: confronted with several possibly entertaining television programs, a typical response is to watch the preferred program most of the time but switch over to check out the other choices during the commercials, just in case something interesting is going on.

The Matching Law

Look more closely at the pigeon in Figure 5.6A and its choice between disk A on a VI 2-min schedule and disk B on a VI 1-min schedule. Can we be more precise about the pigeon’s allocation of time? One way to discover the pigeon’s strategy is simply to let the pigeon peck away for a few minutes, according to its own preference, and then calculate the proportion of time it spent pecking at each disk (Figure 5.6B). In fact, such calculations show that a pigeon will spend about 67% of its time pecking at B and about 33% of its time pecking at A—or about twice as many pecks at B as at A (Herrnstein, 1961).
Note that this 2:1 ratio is identical to the relative rates of reinforcement on the two disks because B is reinforced twice as often as A. On the other hand, if A and B are reinforced equally often (say, both on a VI 1-min schedule), the pigeon will divide its time approximately equally between the two disks. This idea that an organism’s response patterns will closely mimic the relative rates of reinforcement for each possible response is called the matching law of choice behavior.

Of course, even within a lab experiment, a rat or pigeon has more options than just pecking at illuminated disks: it will spend some of its time eating the food it earns, and it can even take a break and spend some of its time grooming, exploring, or napping. Nevertheless, the matching law is a fairly good description of how the animal will allot its time and effort among a set of possible operant responses.

Behavioral Economics

A pigeon confronted with two disks is the simplest possible example of a choice situation. Outside the laboratory, choices are far more complicated. A college student has to divide her allotted studying time among different classes according to how that studying is likely to pay off best, and she has to divide her total time among studying, sleeping, eating, socializing, and so forth. A dieter who is allowed a fixed number of calories per day must decide whether to eat several low-calorie meals or splurge on a bowl of ice cream (and then survive on water and lettuce for the rest of the day).

Behavioral economics is the study of how organisms allocate their time and resources among possible options. For example, a worker who makes $3,000 a month after taxes can distribute this income on rent, food, new clothes, savings, and so on. If she lives in a very expensive apartment, there is less money available for fancy food and new clothes; if she rents a less expensive apartment, she has more money to spend on other things. How does she choose?

Economic theory predicts that each consumer will allocate resources in a way that delivers the maximum “subjective value,” or relative satisfaction, for that person. (In microeconomics, the word utility is used instead of subjective value.) The value is subjective because it differs from person to person: one individual may find much subjective value in an expensive apartment, but another may find more subjective value in having extra money to spend on clothes and food. The particular allocation of resources that provides maximal subjective value to an individual is called the bliss point (Allison, 1983; Timberlake, 1980). We determine an individual’s bliss point simply by recording what that individual chooses to do.

For example, consider Jamie, a college student with a part-time job that pays $100 per week. Assuming that Jamie has no other expenses, he can spend this money to go see live music with his friends (say, $20 admission at a local club) or splurge on his favorite morning coffee (say, $10 for a large caramel latte with double whip from a trendy cafe). Each week he can spend the full $100 to see 5 bands or buy 10 lattes or any other combination of expenditures that adds up to $100. Figure 5.7A shows Jamie’s possible options; each point on the line is a different combination he can purchase with his $100.

Given these options, what does Jamie actually do? Suppose we record his behavior over the semester and find that, most weeks, he buys 6 lattes and sees 2 bands. This point (shown in Figure 5.7A) is Jamie’s bliss point—the distribution of expenditures that (apparently) results in maximum subjective value for this individual: he gets his coffee most mornings, and he also gets to see plenty of live music. Of course, both the curve

and the bliss point can shift if economic conditions change: if the club raises its prices ($50 admission), then Jamie may shift to one evening out and five lattes per week—resulting in the new bliss point shown in Figure 5.7B.

Humans are not the only animals that have to choose how to allocate their time and energy among competing options. College student Jamie has a bliss point that reflects the monetary cost of his different options. Although animals do not use money, they also allocate their behaviors in a way that reflects the “cost” of each option. For example, predators such as sunfish must decide whether to invest energy chasing down a small prey animal like a guppy or wait for a larger one to come along—which could result in a bigger meal for the same energy expenditure. How to choose? One factor is the density of prey. If prey are rare, then the predator should probably chase any meal it sees, small or large. But if prey are plentiful, then the predator might as well skip the guppies and wait for a larger victim—which is just what the sunfish does. In a laboratory tank where large and small fish are plentiful, a sunfish generally only bothers to chase large prey. But when there are only a few fish in the tank, the sunfish will go after both large and small prey—whatever it can find (Warner & Hall, 1974). The sunfish’s bliss point—the allocation of its resources among different classes of prey—changes when environmental conditions change, just as Jamie’s bliss point shifts when economic conditions change.

Similarly, a naive rat in a Skinner box will spend some amount of time grooming, exploring, and sitting quietly—and a very small amount of time pressing the lever. Now, assume each lever press results in delivery of a single pellet of rat chow on an FR 1 schedule. As the rat learns this association (and assuming that the rat is hungry), it will begin to spend more of its time on lever-pressing (and eating) and proportionately less time on other activities because the subjective value of pressing the lever has been increased.
If a second lever is added, and if pressing that new lever results in a preferred food (sugar pellets) on an FR 1 schedule, the animal will tend to press the new lever more often than the old. But if the “cost” of obtaining sugar pellets is then increased, say from an FR 1 schedule to an FR 10 schedule, the rat may switch back to the less preferred but more easily obtainable rat chow—analogous to Jamie’s decision to see live music less often if the cost of admission increases. From a behavioral economics perspective, then, operant conditioning is not so much training an organism to execute a specific behavior as causing the organism to shift its allocation of time and energy among possible behaviors (Baum, 2002, 2004).

Delayed Reinforcement and Self-Control

Another factor that influences how organisms allocate their time and resources is the expected time until the outcome arrives. For example, given a choice between two levers, one that produces a large reward (say, 3 food pellets) and one that produces a small reward (say, 1 pellet), rats reliably respond by pressing on the high-payoff lever. Similarly, given a choice between an envelope containing $500 and an envelope containing $1,000, you would probably choose the larger sum. But what if the choice were between $500 right now and $1,000 mailed to you in 1 year? This choice would be more difficult because—for most people—the subjective value of a reward is progressively reduced (or discounted) the longer that reward is delayed. This phenomenon is known as delay discounting. Put simply, a bird in the hand is worth two in the bush.

For instance, suppose a rat is trained in a Skinner box with two levers: pressing one lever results in delivery of a small, immediate food reinforcement (1 food pellet), whereas pressing on the other lever results in a larger food reinforcement (3 pellets) that arrives after some delay. As shown in Figure 5.8, if the delay is short (e.g., 1 or 2 seconds), a rat will choose the larger reward nearly 100% of the time. But as the delay gets progressively longer, the rat becomes less willing to wait, opting more and more often for the smaller, immediate reward (e.g., Mobini et al., 2002).

The same trade-off occurs in humans: it is easy to convince a student to put off fun in order to study if the exam is coming up tomorrow; it is harder if the exam is not for 5 weeks, even though starting to study early will result in a better grade. The delay between response (early start to studying) and reinforcement (good grade) makes the reinforcement less effective in evoking the response (as demonstrated in Figure 5.4). Similarly, one reason a weight-loss diet is difficult to maintain is that, at each meal, the dieter has to choose between the immediate reward of a dessert and the delayed reward of future weight loss.

Self-control refers to an organism’s willingness to forgo a small immediate reward in favor of a larger future reward. The degree of self-control differs across individuals and across the age span. For example, when asked, “Would you rather receive $500 today or $1,000 in 1 year?” adults in their 60s are highly likely to choose the larger, delayed reward; college students are somewhat less likely to choose the delayed reward; and 12-year-olds almost never choose the delayed reward—preferring immediate gratification (Green, Fry, & Myerson, 1994).

One way of improving an individual’s ability to wait for a reward is to induce the person to make a precommitment—that is, to make a choice...
that is difficult to change later. So, for instance, a student may be more likely to study early in the semester if she joins a weekly study group, in which case she will experience peer pressure to attend the group and to study a little each week. A dieter may be less likely to cheat on his diet if he first empties the kitchen of chips and ice cream, so that when the cravings hit, it will be difficult to sneak some junk food. These precommitments do not make it impossible to get the immediate reward (the student can skip a study meeting, and the dieter can drive to the supermarket to buy ice cream), but they do make it harder to get the immediate reward, and the individual is consequently more likely to stick by an earlier decision in favor of the later, larger reward.

**Reciprocity and Altruism**

As you have now seen, many seemingly complex choice behaviors can be explained in terms of operant conditioning, involving an $S \rightarrow R \rightarrow O$ association. You have learned that people and other animals will choose among possible behaviors in a way that maximizes expected subjective value and that they can show self-control, or the willingness to work now for anticipated future rewards. These behaviors all seem generally consistent with Thorndike’s law of effect (described earlier), which states that actions resulting in desirable outcomes (reinforcers) should increase in frequency and that actions resulting in undesirable outcomes (punishers) typically decrease in frequency.

But there are other behaviors that seem to violate the law of effect. An important example is **altruism**, which occurs when organisms act in ways that benefit another—at the expense of some cost to themselves. Examples of altruism in human behavior would be an individual donating food or money to help victims of a natural disaster and someone waiting in line allowing another to go first.

Other animals appear to show altruistic behavior as well. Among social insects such as some bees and wasps, individuals may sacrifice their own lives to care for and defend the queen and her offspring in the hive (Gadagkar, 2016; Nouvian, Reinhard, & Giurfa, 2016). Among meerkats, a male will “volunteer” to babysit for the day, sacrificing his own opportunity to eat, while others in the group forage for food (Carlson et al., 2006). And dolphins have been observed swimming for hours underneath an injured female dolphin, helping to support her body near the surface so she can continue breathing (Park et al., 2013). These seemingly altruistic actions benefit the recipients—but carry high costs to the actors. Yet the principles of operant conditioning predict that if an action results in punishment, such as loss of time and resources, then the frequency of that action should decrease, and altruistic behaviors should extinguish. Why, then, do these altruistic behaviors persist?

Many apparently altruistic acts occur in the context of kin relationships, as when parents sacrifice time and energy to feed and care for their young, sometimes risking death or injury to protect the young from predators. Among many social animals, such as meerkats and social insects, most of the individuals in a colony are closely related—an extended “family.” In these cases, animals performing altruistic acts sacrifice their own well-being to promote the welfare of their kin. This has been explained in terms of natural selection: an individual’s kin (children, siblings, nieces, and nephews) share many of the same genes, so that acting to help those kin survive into the next generation is actually a way to promote the survival of one’s own genes (Hamilton, 1963). Thus, genes that predispose animals to behave in seemingly altruistic ways toward their kin are genes that are likely to be passed down to future generations, which will in turn also tend to behave in altruistic ways. Much seemingly altruistic behavior among kinship groups may be interpreted this way (Griffin & West, 2003).
However, not all altruistic behavior is easily explained in terms of promoting kin survival. Altruistic acts sometimes occur between “friends” who are not genetically related. For example, genetically unrelated macaque monkeys will sit and groom one another, combing through each other’s fur to find and remove small parasites (Adiseshan, Adiseshan, & Isbell, 2011); unrelated male dolphins may form “alliances” to help guard each other’s female consorts (Connor, 2010); and red-winged blackbird males will help to defend unrelated neighbors’ nests (Olendorf, Getty, & Scribner, 2004). Why would animals behave this way?

One possible explanation for these actions is the principle of **reciprocal altruism**, which states that an organism may donate time or resources to help another in the expectation that the individual will return the favor later on (Trivers, 1971). This is similar to the concept of “tit-for-tat” behavior, which can be roughly restated as “you scratch my back, I’ll scratch yours” (Axelrod & Hamilton, 1981). For example, monkey A grooms monkey B now, in expectation that B will reciprocate later on. If that expectation pays off, it provides delayed reinforcement for monkey A’s original act, which in turn increases the likelihood that A will groom B again in the future. Similarly, humans who lend money or emotional support to a friend may do so in the expectation that someday they themselves might need help, and the friend will return the favor.

In this way, many acts of friendship and cooperation among humans and other animals can be interpreted in the context of operant conditioning:

\[
S^D (\text{friend needs favor}) \rightarrow R (\text{perform favor now}) \rightarrow O (\text{receive future benefits})
\]

Consistent with this interpretation, male red-winged blackbirds will stop defending their neighbors’ nests if those neighbors do not reciprocate (Olendorf et al., 2004), and humans will typically stop behaving generously to friends who fail to reciprocate in some fashion to sustain the friendship. Across a range of species, animals seem to be able to keep track of services given and received over considerable periods of time, and animals that have recently received small acts of “friendship,” such as co-grooming, may be more likely to offer more valuable services, such as defense, in the future (Seyfarth & Cheney, 2012).

But there are other examples of altruism that do not obviously involve expectations of reciprocity, as when a donor gives money to a charity that will redistribute the money to needy individuals who will never know the identity of the donor—and who are thus unable ever to reciprocate directly. Some humans even donate organs to strangers they will never meet, at potential cost to their own health and fitness. Even in the laboratory, when participants are asked to allocate resources between themselves and an anonymous partner, some participants (but not all!) choose to behave altruistically even though they will never meet the partner again (Seymour, Yoshida, & Dolan, 2009).

These selfless acts, undertaken without expectation of direct repayment, are harder to explain in the context of the law of effect. Possibly, the actor receives reinforcement in the form of societal approval of the good deeds, or even an internal feeling of “doing good” that reinforces the altruistic behavior. More recently, researchers have proposed that, at least for animals in social groups, the simple rule of “tit-for-tat” can be expanded to a more general rule of “help anyone if helped by someone” (Taborsky, Frommen, & Riehl, 2016). For instance, a male meerkat may volunteer to babysit today, protecting his own offspring along with the rest of the infants, in the general expectation that other males in the group will take their turns on subsequent days. Similarly, a person may help a stranger today in expectation that, in the future, someone else (or society at large) will come to his aid in return.
Of course, operant conditioning cannot easily explain all instances of altruism; in humans, cultural, philosophical, and religious factors also play a role. Nevertheless, altruism is an example of a very complex behavioral phenomenon that can at least partially be interpreted via the principles of operant conditioning and delayed reinforcement.

**Interim Summary**

- In operant conditioning, organisms learn to make responses under particular conditions in order to obtain or avoid outcomes: Discriminative stimulus $S^D \rightarrow$ Response $R \rightarrow$ Outcome $O$.
- In operant conditioning, the outcome (reinforcement or punishment) occurs only if the organism makes the response. In classical conditioning, by contrast, the unconditioned stimulus (US) occurs whether or not the organism makes a conditioned response (CR).
- Discriminative stimuli signal to the organism whether a particular response will result in a particular outcome.
- A reinforcer is an outcome that an organism will work to obtain; a punisher is an outcome that an organism will work to avoid. While punishment can be effective in eliminating an undesired response, it leads to more varied behavior and can be undermined by discriminative stimuli that encourage cheating, by concurrent reinforcement, or by weakness of the initial punisher. Another approach to eliminating unwanted behavior is differential reinforcement of alternative behaviors (DRA).
- Complex responses can be trained by shaping, in which progressive approximations to the desired response are reinforced, as well as by chaining, in which organisms are gradually trained to execute a sequence of responses.
- The four basic types of operant paradigm are positive reinforcement, negative reinforcement, positive punishment, and negative punishment. The words *positive* and *negative* denote whether the outcome is added to or subtracted from the environment; *reinforcement* and *punishment* denote whether the response increases or decreases in frequency as a result of learning.
- Schedules of reinforcement define whether the outcome $O$ follows every response $R$, is available after some (fixed or variable) number of responses, or is available only after some (fixed or variable) time interval.
- When different responses are reinforced under VI schedules, the matching law predicts that organisms will allocate time among those responses based on the relative rates of reinforcement for each response.
- Behavioral economics is the study of how organisms choose to allocate their time and resources among various responses that result in different outcomes. The bliss point is the particular allocation of resources that provides maximal subjective value to an individual.
- Altruism refers to actions that are costly to the actor but that benefit another. Altruism toward one’s kin can partially be explained in terms of natural selection. Altruism among “friends” may partially be explained as operant conditioning in terms of a tit-for-tat model.
5.2 Brain Substrates

The previous section defines operant conditioning as learning an $S^D \rightarrow R \rightarrow O$ association. In studying such associations, neuroscientists are discovering that the parts of the brain that link stimuli with responses ($S^D \rightarrow R$ learning) are different from the parts of the brain that learn about the expected outcomes of those responses ($R \rightarrow O$). While many brain areas play roles in these processes, two key areas are the dorsal striatum, which appears to be particularly important for $S^D \rightarrow R$ learning, and the orbitofrontal cortex, which appears to be important for learning about expected outcomes. Still other brain areas may help us evaluate whether those outcomes are reinforcers or punishers.

The Dorsal Striatum and Stimulus–Response ($S^D \rightarrow R$) Learning

Voluntary motor responses occur when neurons in the motor cortex send messages to motor neurons in the muscles that control movements. The motor cortex receives its primary inputs from cortical areas that process sensory information, such as the visual cortex (V1) and the somatosensory cortex (S1), which you saw back in Figure 2.8. It also receives inputs from the frontal cortex. When you see a book, this visual stimulus is registered by your visual cortex. If you decide to pick up the book, this “decision” is made in your frontal cortex. Signals from both the visual cortex and the frontal cortex then travel to your motor cortex, which integrates these signals and produces the appropriate instructions, causing your arm and hand to move in a coordinated way as you pick up the book.

Information from the sensory cortex to the motor cortex can also travel by an indirect route, through the basal ganglia. The basal ganglia are a collection of ganglia—that is, clusters of neurons—that lie under the cortex, on either side of the thalamus. One part of the basal ganglia is the dorsal striatum, shown in Figure 5.9, which can be further subdivided into the caudate nucleus and the putamen. The dorsal striatum receives highly processed stimulus information from the sensory cortex and projects it to the motor cortex, which produces a behavioral response.

The dorsal striatum plays a critical role in operant conditioning, specifically the ability to learn $S^D \rightarrow R$ associations based on feedback about reinforcement and punishment (McDonald & White, 1994; O’Doherty et al., 2004). Rats with lesions of the dorsal striatum can learn simple operant responses (e.g., when placed in a Skinner box, they can learn to lever-press R to obtain food O). But if discriminative stimuli are added (e.g., if lever-press R is reinforced only in the presence of a light $S^D$), then the lesioned rats are markedly impaired (Featherstone & McDonald, 2004). In humans, too, individuals with damage or disruption to the striatum due to Parkinson’s disease or Huntington’s disease show deficits in the ability to associate a discriminative stimulus with a correct response (Ashby & Waldron, 2000; Robbins, 1996).
SD → R associations that depend on the dorsal striatum tend to be relatively automatic or habitual (Balleine, Daw, & O’Doherty, 2008). Remember the well-trained rats, discussed earlier in the chapter, that would run right through a pile of food on their way to a goal box in the maze? That behavior probably reflects SD → R learning in the striatum, making the maze-running automatic even when other behaviors (such as pausing to eat) would have resulted in reward. In this case, running is based on a history of learning that the response resulted in desirable outcomes; but after a long period of training, the response is performed even though the outcome is no longer contingent on that action.

The Orbitofrontal Cortex and Learning to Predict Outcomes

SD → R learning is, of course, only part of the picture in operant conditioning. Organisms learn to predict that particular responses R (in the presence of SD) will result in particular outcomes O. For example, in Section 5.1, you read about the negative contrast effect: monkeys may shriek in annoyance if their response earns them a less preferred food than the one they expected, and trick-or-treaters may feel cheated if they receive raisins rather than the expected candy. Such results show that organisms make responses in anticipation of particular outcomes.

Several brain areas appear to be involved in learning to predict the outcomes of behavior. Among these are parts of the prefrontal cortex, including the orbitofrontal cortex, which lies at the underside of the front of the brain in primates (see Figure 5.9) and which appears to contribute to goal-directed behavior by representing predicted outcomes (Schoenbaum, Roesch, Stalnaker, & Takahashi, 2009; Tanaka, Balleine, & O’Doherty, 2008). The orbitofrontal cortex receives inputs that convey the full range of sensory modalities (sight, touch, sound, etc.), as well as visceral sensations (including hunger and thirst), allowing this brain area to integrate many types of information. Outputs from the orbitofrontal cortex travel to brain areas, including the striatum, where they can help determine which motor responses are executed.

Predicting Specific Outcomes

Some of the strongest evidence that the orbitofrontal cortex plays a role in predicting the outcomes of responses comes from neuronal recordings. For example, thirsty rats can be trained on a discrimination task in which the discriminative stimuli are two odors, the response R is to poke the nose into a nearby water cup, and the two possible outcomes are a tasty sucrose solution or a bitter quinine solution. A short delay (typically less than a second) is introduced between the response and the outcome, during which period a trained animal is “expecting” the outcome. During this delay, some neurons in the orbitofrontal cortex fire differently, depending on whether the animal expects a reward or a punisher (Schoenbaum, Chiba, & Gallagher, 1998). Figure 5.10A illustrates the firing patterns of one neuron in the orbitofrontal cortex of a rat learning such a task. This particular neuron fires strongly if the rat has just made a (mistaken) response to odor 2 and is expecting quinine, but it fires less strongly if the rat has just made a response to odor 1 and is expecting sucrose. Therefore, this neuron appears to code expectation of the punisher rather than the reward. If the contingencies are reversed, such that odor 1 now predicts quinine and odor 2 now predicts sucrose, those neurons often alter their responses to reflect the new contingencies (Stalnaker, Franz, Singh, & Schoenbaum, 2007).

Neurons in the orbitofrontal cortex do not merely learn whether to expect reinforcement or punishment; they apparently even code the actual identity of the expected
outcome. Thus, monkeys can be trained to differentiate a set of pictures that predict whether the upcoming reward will be grape juice or orange juice. Figure 5.10B shows the responses of a single neuron that became active whenever a stimulus that predicted grape juice was presented but not when pictures predicting orange juice were presented (Tremblay & Schultz, 1999). This same neuron also fired during the actual delivery of grape juice but not during the delivery of orange juice. Given their ability to encode specific predicted outcomes, orbitofrontal cortex neurons play an important role in helping us select between potential actions based on their expected consequences. Recall college student Jamie, who could spend his weekly income by distributing it among choices such as music and lattes. Possibly, neurons in Jamie’s orbitofrontal cortex were helping him evaluate the potential outcomes of his actions so that he could act to maximize the expected subjective value.

Delayed Reinforcement and Self-Control

In addition to processing the identity of expected rewards, the orbitofrontal cortex appears to evaluate the expected timing of those rewards. Remember the rats in Figure 5.8 that could choose between pressing one lever to obtain a small immediate food reward or pressing another lever for a larger but delayed reward? The data are replotted in Figure 5.11A (“control”). As you saw before, if the larger reward is due after a short delay (e.g., less than 2 seconds), healthy (control) rats will almost always choose to wait for that larger reward; but as the delay for the larger reward increases (e.g., 30 seconds), the rats become increasingly more likely to choose the smaller, immediate reward. Presumably, the rats are making an internal calculation of subjective value, comparing the two expected outcomes. The orbitofrontal cortex appears to be important for making these assessments of subjective value. Figure 5.11A (“OFC lesion”) shows that rats with lesions of the orbitofrontal cortex will still choose the delayed reward if the delay is short; however, as the delay increases, their willingness to wait plummets (Mobini et al., 2002). In effect, these lesioned rats show decreased self-control—less willingness to work for an expected future payoff.
Humans show a similar effect, as illustrated in Figure 5.11B. Healthy humans, given a choice between a small immediate reward (say, $20 now) or a larger reward after a short delay (say, $40 in 30 days), will usually prefer the larger reward if the delay is short (e.g., a few seconds), but as the delay increases, they will increasingly prefer the immediate reward. People with damage to the orbitofrontal cortex also prefer the larger reward if the delay is short; but as the delay increases, their ability to delay gratification plummets (Sellitto, Ciaramelli, & de Pellegrino, 2010). In effect, they become like college students who will study the night before an exam but are unwilling to start studying a month or a week earlier in anticipation that the effort now will pay off later.

**Mechanisms of Reinforcement Signaling in the Brain**

Operant conditioning involves learning that responses (R) in the presence of certain stimuli (S\textsuperscript{D}) will result in particular outcomes (O). You have just read that neurons in the orbitofrontal cortex help encode the identity of specific outcomes (e.g., grape juice versus orange juice). If an outcome is reinforcing, the S\textsuperscript{D} → R association should be strengthened, increasing the likelihood that S\textsuperscript{D} evokes R in the future; if it is a punisher, the association should be weakened, decreasing the likelihood of R. But how does the brain determine whether an outcome should be interpreted as a reinforcer or a punisher?
“Wanting” and “Liking” in the Brain

In 1954, James Olds was experimenting with delivering electrical stimulation to the rat brain. He inserted an electrode into an area that researchers now believe to have been the lateral hypothalamus. Olds waited until the rat wandered into a certain corner of the experimental chamber, and then he applied a brief electrical current. A few minutes later, the rat came back to the same corner, and Olds provided a second stimulation. The rat caught on quickly and began to loiter in that corner of the chamber, apparently hoping for more electrical stimulation (Olds, 1955). Thus, electrical stimulation to this area of the brain seemed to be acting much like a reinforcer, increasing the probability of certain responses (in this case, hanging around the correct location).

Olds was intrigued, to say the least. He rigged a Skinner box so that the rats could press a lever to turn on the electrical stimulation. The rats were soon lever-pressing at a furious rate—as much as 700 presses per hour (Olds, 1958). If allowed, rats would press the lever continuously for up to 48 hours, until they collapsed from physical exhaustion! Given a choice between electrical stimulation and food, the rats would literally starve themselves, preferring the stimulation (Routtenberg & Lindy, 1965).

Later studies identified that rats would work for electrical stimulation in several brain areas, in particular the ventral tegmental area (VTA), a small region in the midbrain of rats, humans, and other mammals (Figure 5.12). The electrodes in Olds’s original studies were probably stimulating hypothalamic neurons with axons that project to the VTA, indirectly activating this area. Because VTA stimulation was such a powerful reinforcer, some researchers inferred that the rats “liked” the stimulation, and the VTA and other areas of the brain where electrical stimulation was effective became informally known as “pleasure centers.”

However, the idea of “pleasure centers” is an oversimplification. For one thing, rats pressing levers for electrical brain stimulation do not tend to act as if they are enjoying it; rather, they tend to become agitated and may bite the lever instead of simply pressing it, or they may show other behaviors, such as fighting, scratching the walls, or shredding the nesting material. This is more like the behavior of an excited animal than one who is enjoying food. Skinner, of course, would caution that we can’t infer what an animal might be feeling just by watching its behaviors. Nevertheless, some researchers suggested that electrical brain stimulation causes not pleasure but rather excitement or anticipation of reinforcement—much like the anticipation we experience when expecting a good meal or a big present (Flynn, 1972).

Currently, many researchers believe that we have separate brain systems for signaling hedonic value—meaning the subjective “goodness” of a reinforcer, or how much we “like” it—that are distinct from those signaling motivational value—meaning how much we “want” a reinforcer and how hard we are willing to work to obtain it. No matter how much we may “like” chocolate cake, most of us will not be very motivated to obtain more if we have just eaten three slices; similarly, Olds’s rats doubtless still “liked” food and rest, but they were more motivated to obtain electric brain stimulation, even when starving and exhausted. As these examples demonstrate, provision of a “liked” reinforcer is not enough to evoke responding. Only when “wanting” and
“liking” signals are both present will the arrival of the reinforcer evoke responding and strengthen the $S^D \rightarrow R$ association.

**Dopamine: How the Brain Signals “Wanting”?**

The neurotransmitter dopamine is produced by neurons in the ventral tegmental area (VTA), a part of the midbrain that projects to the frontal cortex (among other places); dopamine is also produced in the nearby substantia nigra pars compacta (SNc), a part of the basal ganglia that projects to the striatum (Figure 5.12). As you read earlier in this chapter, the dorsal striatum is an important site of $S^D \rightarrow R$ association, and the orbitofrontal cortex is important for learning about predicted outcomes, so dopaminergic neurons in the VTA/SNc are a good place to start looking at how the brain signals motivational value.

In rats, dopamine release from the VTA/SNc is triggered by encounters with food, sex, drugs of abuse, and secondary reinforcers. In humans, PET and fMRI studies have shown that presentation of juice, cocaine, money, humor, and even video games causes heightened activity in dopamine target sites such as the striatum (Berridge & Robinson, 1998; Knutson, Fong, Adams, Varner, & Hommer, 2001; Mobbs, Greicius, Abdel-Azim, Menon, & Reiss, 2003). Even in invertebrates, such as the sea slug *Aplysia*, dopamine is released in conjunction with positive reinforcement during operant conditioning (Brembs, 2003; Nargeot, Baxter, Patterson, & Byrne, 1999).

Many researchers believe that dopamine does not simply signal hedonic value, or “liking.” For example, Parkinson’s disease is caused by the progressive death of dopamine-producing neurons, many of which project to the striatum. But when patients with Parkinson’s disease are asked to rate the perceived pleasantness of sweet and salty tastes, their ratings are the same as those of healthy people. Apparently, the dopamine reduction in these patients causes no loss of the ability to “like” pleasurable stimuli (Travers et al., 1993).

Similar results are obtained from nonhuman animals. Researchers can’t simply ask rats to rate the perceived pleasantness of different tastes. But just as a human child given a taste of sugar may lick her lips, a monkey or rat given sugared water will show rhythmic movements of the mouth and protrusion of the tongue. This is sometimes called the hedonic, or “yum,” reaction. On the other hand, a bitter taste produces a different cluster of responses: mouth gapes, nose wrinkling, shakes of the head, and wiping of the face with paws or hands—an aversive, or “ugh,” reaction that appears remarkably similar in humans, monkeys, and rats. Researchers can use the “yum” and “ugh” reactions to infer degree of liking in nonhuman animals (and in human infants). Rats given injections of a drug that destroys dopaminergic neurons...
neurons exhibit hedonic ("yum") and aversive ("ugh") responses that are just as strong as or stronger than those of control rats (Berridge & Robinson, 1998). This suggests that rats with damaged dopamine systems continue to "like" food just as much as control rats do. What seems to change is their willingness to work for it.

The **incentive salience hypothesis** of dopamine function states that one role of dopamine in operant conditioning is to signal how much the animal "wants" a particular outcome—that is, how motivated the animal is to work for it. According to this hypothesis, the incentive salience of food and other reinforcers—that is, their ability to attract attention and motivate responding—is reduced in dopamine-depleted animals (Berridge, 1996, 2007; Berridge & Robinson, 1998).

A good example of the incentive salience concept is demonstrated in behavioral economics experiments. For instance, most healthy rats prefer sugar pellets to rat chow. If they have to work for the pellets by lever-pressing but chow is freely available, they will typically choose to work for the pellets (Salamone, Arizzi, Sandoval, Cervone, & Aberman, 2002). Rats given a dopamine antagonist also prefer sugar pellets to rat chow, if both are freely available. But if those same rats have to work for the sugar pellets by lever-pressing, they work much less than control rats do (Figure 5.13A) and mostly settle for eating the free chow instead (Figure 5.13B). In other words, given a choice between competing alternatives, normal animals will tend to choose their preferred reinforcer, even at the cost of a little extra work. In contrast, dopamine-depleted animals are still perfectly willing to eat the preferred food if it is placed in front of them, but they are not willing to work hard to earn it.

In an even more extreme case, mice that have been genetically engineered to be completely unable to produce dopamine will not seek food at all, and they generally starve to death by about 20 to 30 days of age—even if pellets are placed directly in front of them (Palmiter, 2008). However, if the food is placed in their mouths, these animals will chew and swallow and even exhibit "yum" responses, indicating that they still "like" food and can consume it; they just lack the motivation to obtain it. These mice can be "rescued"...
by infecting cells in the striatum with a recombinant virus that allows the cells to produce and release dopamine; afterward, the mice eat enough normal chow to maintain body weight without further intervention.

Dopamine seems to affect incentive salience in humans, too. For example, in one study, humans addicted to cocaine were given a drug that increases brain dopamine levels; these participants reported an increased craving for cocaine but no increase in the self-reported “high” from cocaine (Haney, Foltin, & Fischman, 1998). Thus, stimulating the dopamine system increases “wanting” but not “liking” of cocaine.

The dopamine system can also be stimulated naturally by exposing the organism to a stimulus that has had its “temptation power” boosted by previous association with reinforcement (Berridge, 2012). For instance, the sight of chocolate can stimulate intense desire in chocolate lovers, even if they are not particularly hungry; similarly, a cigarette smoker who is sincerely trying to quit may experience an overwhelming craving if he enters a room where he can see and smell others smoking. For this reason, many smokers and drug addicts who wish to quit try to avoid environments where they are likely to encounter people using the addictive substance.

In addition to dopamine’s role in “wanting,” there is also considerable evidence that dopamine plays a key role in monitoring reward prediction error, meaning the degree to which an outcome is unexpected (Schultz, 1998, 2015). Reward prediction error information is critical when an organism is learning how to predict (and act to obtain) future reward. Remember the experiment, described in Figure 5.10B, in which the monkey learned that picture 1 predicted upcoming grape juice, while picture 2 predicted upcoming orange juice? Imagine that now picture 1 is followed by orange juice (surprise!). When the unexpected outcome occurs, the monkey’s dopamine neurons spike sharply—in effect, signaling that something new and unexpected is happening. This signal travels to the orbitofrontal cortex (along with other brain areas), facilitating updates to $S^{D} \rightarrow R \rightarrow O$ associations to reflect the new reality, so that the monkey can continue to predict which responses will achieve its preferred outcomes.

Endogenous Opioids: How the Brain Signals “Liking”?

If dopamine signals “wanting,” then what signals “liking” in the brain? Probably the best-studied candidate is the opioid system. Opiate receptors in the brain were discovered quite by accident in the 1970s by researchers trying to figure out how heroin and morphine work. Heroin and morphine belong to a class of drugs called opiates, which bind to a class of neuronal receptors called opiate receptors. Rather than assume that the brain evolved special receptors to respond to heroin and morphine, researchers suspected there might be naturally occurring brain chemicals that also activate the opiate receptors—and indeed there are. The endogenous opioids are naturally occurring neurotransmitter-like substances that have many of the same effects as opiate drugs. (The word endogenous means “originating on the inside”; opoid means “opiate-like.”) Endogenous opioids are distributed throughout the central nervous system, and when released into the body, they have a wide range of effects, such as reducing the normal perception of pain and producing feelings of euphoria.

Although there is still a great deal to be learned about the endogenous opioids, many researchers believe these substances may mediate hedonic value, or “liking.” If so, the reason that heroin and morphine are so intensely pleasurable is that they happen to activate the same brain receptors as the endogenous opioids.

For example, morphine makes sweet food taste sweeter and bitter food taste less bitter (Rideout & Parker, 1996). It can also make pain feel less painful; morphine is used
medically for patients who are enduring extreme, long-term pain (in cases where the benefits of relieving suffering outweigh the risks of morphine addiction). These patients usually report that they still feel the pain but that it does not trouble them as much as it did before.

Endogenous opioids are released in response to primary reinforcers, such as food, water, and sex, and they may be released in response to secondary reinforcers and pleasurable behaviors, too (Le Merrer, Becker, Befort, & Kieffer, 2009). Differences in the amount of endogenous opioid released and in the types of opiate receptors they activate may help determine an organism’s preference for one reinforcer over another (Le Merrer et al., 2009). This would contribute to effects such as the ones shown in Figure 5.3, where infants will suck harder to obtain sweetened water even though plain water satisfies thirst just as effectively. Just like infants, rats normally prefer sweetened water to plain water, but rats given the opioid antagonist naloxone choose the sweetened water much less often than do control rats (Hayward, Schaich-Borg, Pintar, & Low, 2006)—as if they no longer “like” the sweetness as much.

How Do “Wanting” and “Liking” Interact?

If dopamine signals “wanting” and the endogenous opioids signal “liking,” how might these two brain systems interact to influence behavior? The answer is not yet clear. One possibility is that endogenous opioids may modulate dopamine release. For example, some neurons in the VTA have opiate receptors on their dendrites that, when activated, could affect those neurons’ normal tendency to release dopamine. In this way, the endogenous opioids could signal “liking,” which in turn could affect the VTA’s ability to signal information about “wanting.” But other studies have suggested that different subpopulations of dopamine neurons might exist, conveying salience (“wanting”) and valence (“liking”) separately (Matsumoto & Hikosaka, 2009). The picture is complicated because some drugs, such as heroin, may manipulate both pathways—activating the “liking” system to produce a pleasurable high, while also activating the “wanting” system to produce a craving for more of the drug and the high.

Punishment Signaling in the Brain

If the dopamine and opioid systems help code “liking” (hedonic value) and “wanting” (motivational value) of reinforcers, then what codes the aversive value of punishers? So far, there does not appear to be just one, unique “pain center” in the brain. Rather, both physical and emotional pain can activate multiple pathways and systems in the brain.

Physical pain often begins in the skin or musculature, where specific receptors called nociceptors respond to intense pressure, heat, or other stimulation that can cause damage. Messages from these receptors pass through the brainstem and thalamus to reach somatosensory areas in the cortex, such as the primary somatosensory cortex (S1; introduced in Chapter 2 and illustrated in Figure 2.8). Brain-imaging studies have shown that the more intense the pain, the more activity in S1. When you shower, for example, as the water gets hotter, you will have more activity in S1. But although S1 encodes the physical location and intensity of pain, it does not encode how bad the pain “feels”—the affective component of pain. If you have spent all day freezing outside in the snow, standing under that same very hot shower may actually feel good rather than painful. Similarly, a man swallowing wasabi-flavored snacks may gasp for breath and wipe away tears—and then reach for another handful to do it again. Clearly, not all intense stimuli are aversive, and not all aversive stimuli cause physical pain: disgusting smells, loud discordant sounds, and social rejection can all be highly aversive, even though no physical pain occurs.
The Insula: How Much Does It Hurt?

So how does the brain decide whether a particular stimulus is aversive? Several brain areas have been implicated, including the insular cortex, or insula, shown in Figure 5.14. The insular cortex is located in the deep fold separating the temporal lobe from the parietal and frontal lobes, and it is important for our conscious awareness of our own bodies and emotional states. One subregion of the insular cortex, the dorsal posterior insula, plays a role in perception of physical pain, as well as in other negative emotional states, such as hunger, anger, and disgust (Chang, Yarkoni, Khaw, & Sanfey, 2013; Naqvi & Bechara, 2009). For instance, functional neuroimaging has shown that the dorsal posterior insula is more active when research participants experience painful heat or cold (Craig, 2003), as well as when they experience social rejection—such as being excluded by other players in an online video game (Eisenberger, Lieberman, & Williams, 2003) or when viewing pictures of an ex-partner after an unwanted breakup (Kross, Berman, Mischel, Smith, & Wager, 2011). The degree of activation appears to be roughly proportional to the intensity of the punisher. So, for example, in a study where errors could be punished by loss of 5 cents or 50 cents, the insula showed more activity after a larger loss (Hester, Murphy, Brown, & Skilleter, 2010).

Thus, just as the opioid system may signal pleasantness, or “liking,” the insula may help the brain determine unpleasantness, or “disliking.” In fact, when the insula is damaged, learning to avoid unpleasant outcomes is impaired. One study trained participants to play a simple computer game and found that people with damage to the insula could learn how to respond in order to obtain reward (point gain), but they were impaired at learning to avoid punishment (point loss) compared with people with brain damage that spared the insula (Palminteri et al., 2012).

**FIGURE 5.14 The insular cortex (insula) and dorsal anterior cingulate cortex (dACC)** The insula, which lies buried in the fold separating the temporal lobe from the parietal and frontal lobes, is implicated in conscious awareness and also plays a role in signaling the aversive value of stimuli. (The figure illustrates the surface of the frontal and temporal lobes being “pulled back” by metal clips to reveal the insula underneath in purple.) The dACC (shown in orange), which lies on the inner, or medial, surface of the prefrontal cortex, may play a role in the motivational value of punishers, helping select the actions we take in response.
The Dorsal Anterior Cingulate Cortex: What to Do About It?

Once we have established that a stimulus is subjectively painful, the next step is to decide whether to do something about it. The dorsal anterior cingulate cortex (abbreviated dACC), which lies on the inner, or medial, surface of the prefrontal cortex (Figure 5.14), has been implicated in the motivational value of pain—the degree to which pain can drive changes in behavior (Craig, 2003). For example, in one study, participants played a game in which they could win or lose money. During the game, neurons in the dACC responded both to errors that resulted in outright punishment and to errors that merely resulted in no reward—but there was more activity in the former case (Simões-Franklin, Hester, Shpaner, Foxe, & Garavan, 2010). Presumably, the worse the consequences of an error, the greater the motivation to change behavior.

The dACC also shows increased activation when study participants unexpectedly receive a reduced reward (Bush et al., 2002; Williams, Bush, Rauch, Cosgrove, & Eskandar, 2004), and the activity level is predictive of whether participants actually change their response (Williams et al., 2004). Remember the phenomenon of negative contrast, in which monkeys and children refuse to work for a reward that is smaller than the one they expect? In effect, the smaller-than-expected reward is functioning as a punisher, leading to decreased responding. It is possible that the dACC is recognizing this negative contrast and signaling reduced motivation to work for the disappointing reward.

Thus, just as the brain has multiple systems for signaling the hedonic value and motivational value of reinforcers—“liking” via the opioid system and “wanting” via the dopamine system—the brain may also have multiple systems to signal the aversive value and motivational value of punishers, via brain areas such as the insula and dACC. However, much remains to be learned about how we process and respond to punishers in the brain.

Interim Summary

■ The dorsal striatum is an important brain substrate for storing stimulus–response (SD → R) associations; these SD → R associations may become relatively automatic and habitual.

■ The orbitofrontal cortex may be an important brain substrate for storing response–outcome (R → O) associations, including both specific identity of outcomes and expected time until those outcomes arrive.

■ Reinforcers and punishers may activate neurons in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc), which project dopamine to the dorsal striatum, frontal cortex, and elsewhere. Interrupting these pathways, with lesions or drugs, disrupts operant conditioning.

■ The incentive salience hypothesis suggests that dopamine modulates “wanting” rather than “liking,” determining how hard an organism is willing to work for a reinforcement. Dopamine may also signal “prediction error,” meaning the degree to which outcomes are unexpected or surprising.

■ The endogenous opioids, which are mimicked by many highly addictive drugs, may signal the hedonic value (“liking”) of reinforcers.

■ Parts of the insula may help us determine subjective “disliking” of painful physiological and psychological stimuli. The dorsal anterior cingulate cortex (dACC) may help determine the motivational value of punishers, which is used to guide changes in behavioral responding.
5.3 Clinical Perspectives

Through the brain’s reinforcement system, animals are hardwired to seek and obtain the things they need for survival (food, water, sleep, etc.) and to avoid the things that threaten survival (pain, sickness, predators, etc.). Unfortunately, this powerful reinforcement system can go awry. As an example, consider the pleasure we feel when we eat fatty food, which ensures that we are sufficiently motivated to repeat the experience. The human brain evolved millennia ago, when our ancestors had to forage for food and could never be sure when they would find their next meal. Fat could be stored in the body and used for energy later, when food was scarce. Under these conditions, seeking out fatty foods was a good strategy for survival. In twenty-first-century America, however, food is easier for most of us to obtain, but our biological drives have not changed, and many—still motivated to obtain the taste of fatty foods—have become dangerously overweight.

Drug addiction is another way in which the reinforcement system can malfunction (or, rather, function only too well). Chapter 4 explains how classical conditioning can contribute to drug addiction, but another large piece of the addiction puzzle is operant conditioning: learned responding to obtain a particular kind of reinforcement. Insights from operant conditioning studies may deepen our understanding of addiction and lead to more effective treatments.

Drug Addiction

We all know people who are “addicted” to their morning cup of coffee, their afternoon chocolate bar, or even their favorite television show. Such people may experience intense cravings and even withdrawal symptoms if forced to go without. For instance, someone who drinks strong coffee every morning may experience a pleasurable alertness after she indulges; she may experience crankiness, sleepiness, headaches, or difficulty paying attention if she goes without; and she may have trouble kicking the habit if she decides to give it up. However, unless the coffee is causing harmful medical consequences or is otherwise interfering with her normal life (for example, if she failed to pay the rent because she’s spending all her income at Starbucks), this would not be considered a serious addiction.

In contrast, pathological addiction is a strong habit (or compulsion) that is maintained despite known harmful consequences (Everitt & Robbins, 2016; Grant, Potenza, Weinstein, & Gorelick, 2010). For instance, a person will be diagnosed as pathologically addicted to cocaine if she is unable to quit, suffers debilitating withdrawal symptoms between highs, and is obsessed with obtaining her next hit of the drug, to the point where she starts neglecting other aspects of her life—such as her family and her job—because nothing else is as important to her as cocaine. Alcohol can drive pathological addiction in the same way: an alcoholic may be unable to give up drinking, even though it has cost him his job, his health, and his family. Similarly, a rat that starves rather than cease lever-pressing to obtain electrical brain stimulation could be considered pathologically addicted to the stimulation.

Remember the college student Jamie who had to choose how to allocate his resources between going clubbing and purchasing expensive coffee (Figure 5.7)? We can think of a person with addiction as having similar choices for allocating resources (time, money, effort) and increasingly choosing the addictive substance at the expense of all other options.
Many individuals with pathological addictions want to quit, and they try very hard to overcome their addictions. Unfortunately, there are several processes working against them. Addiction may involve not only seeking the “high” but also avoiding the adverse effects of withdrawal from the drug. In a sense, the high provides a positive reinforcement, and the avoidance of withdrawal symptoms provides a negative reinforcement—and both processes reinforce the drug-taking responses.

Many highly addictive drugs are opiates, meaning that they target opiate receptors in the brain. Heroin and morphine are two examples of opiate drugs. Other commonly abused drugs, including amphetamines and cocaine, work by increasing brain dopamine levels. As explained in Chapter 2, neurons communicate when a presynaptic (or sending) neuron releases molecules of a neurotransmitter into the synapse, and these neurotransmitter molecules activate receptors on the postsynaptic (or receiving) neuron (Figure 5.15A). Amphetamine causes dopaminergic neurons to release higher levels of dopamine. Cocaine works by blocking dopamine reuptake, and dopamine remains in the synapse longer before being reabsorbed. In both cases, the effect is to increase the amount of dopamine available to activate the postsynaptic neuron (Figure 5.15B).

Both amphetamine and cocaine can be used as reinforcers. Thus, for example, rats and mice will learn to lever-press vigorously for injections of amphetamine or cocaine (McBride, Murphy, & Ikemoto, 1999). Unfortunately, as Chapter 4 points out, long-term drug use can cause physiological changes in the synapse, such that ever larger doses of the drug are needed to get the same effect.

One interesting aspect of cocaine and amphetamine use is that, although “liking” seems to be critical in the early stages of drug use, people with long-term addictions often

---

**FIGURE 5.15** The effects of amphetamine and cocaine on dopaminergic neurons  
(A) A presynaptic dopamine-producing neuron releases dopamine into the synapse (1). These molecules activate dopamine receptors on the postsynaptic neuron (2). Unused molecules are broken down (3) and taken back into the presynaptic neuron, a process called reuptake (4).  
(B) Amphetamine works by causing dopaminergic neurons to make and release more dopamine (1). Cocaine works by blocking the reuptake of unused dopamine molecules (4). Both drugs thus increase the amount of dopamine in the synapse, increasing the chance that dopamine molecules will activate receptors on the postsynaptic neuron (2).
report that they no longer experience an appreciable high from the drug but crave it anyway—as if their “wanting” system has disconnected from their “liking” system and has run amok. In fact, individuals who have taken a dose of cocaine or amphetamine do not typically report feeling happy or pleasured; rather, they feel aroused or excited. These facts fit nicely with the incentive salience hypothesis, which proposes that dopamine is involved in “wanting” but not necessarily in “liking.”

A number of studies have shown that the insula is among the brain areas activated when drug abusers experience cravings; this has been shown for cocaine, alcohol, opiates, and nicotine. In addition, several intriguing studies indicate that for cigarette smokers who happened to suffer stroke damage to the insula, their addiction to cigarettes was practically eliminated (Gaznick, Tranel, McNutt, & Bechara, 2014; Naqvi, Rudrauf, Damasio, & Bechara, 2007; Suñer-Soler et al., 2012). Similarly, animals that have been addicted to amphetamine and are then administered drugs to temporarily inactivate the insula show disruption of their drug-seeking behavior (Contreras, Ceric, & Torrealba, 2007). These studies and others like them suggest that the representation of negative feelings (cravings and withdrawal) by the insula helps maintain the addiction.

In summary, drug addiction is currently thought to result from several factors (Robinson & Berridge, 2000), including positive reinforcement (the pleasurable high), negative reinforcement (withdrawal avoidance), and malfunction of the dopaminergic “wanting” system (the craving). Of course, the relative role of each of these factors varies among individuals, as a function of genetics, personality, and experience. Some people become strongly addicted to drugs after a single use, some become addicted over time, and some can use drugs over a long period without developing pathological addictions (Lüscher, 2016). As yet, there is no way to predict whether a particular individual who starts using a particular drug will become addicted, and it is also impossible to know how difficult it will be for a particular individual to break free of addiction.

Behavioral Addiction

For some people, behaviors such as skydiving or winning at gambling can provide highs that are just as reinforcing as drugs—and just as addicting. Behavioral addictions are addictions to behaviors (as opposed to drugs) that produce reinforcements or highs, as well as cravings and withdrawal symptoms when the behavior is prevented.

Perhaps the most widely agreed-upon example of a behavioral addiction is compulsive gambling. Many people gamble occasionally, buying a weekly lottery ticket, playing online poker once in a while, or spending their vacations in Las Vegas or Atlantic City. These people pay their money, have their fun, and then walk away. But other people get hooked: they start to gamble more and more often, risking progressively larger sums of money, until nothing is as important as the chance to gamble. Up to about 1.6% of the general population may suffer from compulsive gambling at some point in their lifetime, and the problem may be even more widespread in particular groups, such as African Americans and college students (Potenza, Kosten, & Rounsaville, 2001).

**behavioral addiction**

Pathological addiction to a behavior (rather than to a pharmacological substance).
Skinner suggested that one reason gambling is so seductive is that it is reinforced on a VR schedule: you can never be sure when the next big payoff will come, which makes it easy to talk yourself into playing just one more time, then just one more time . . . (Skinner, 1953). Each big win provides a powerful reinforcement of the gambling behavior.

Gambling is not the only behavior that can affect the brain’s reinforcement system and become addictive. Other behavioral addictions include compulsive eating, sex addiction, compulsive exercising, kleptomania, compulsive shopping—the list goes on and on. In each of these cases, the addicted person experiences a high from the behavior, followed by craving for another high, and withdrawal symptoms if the behavior is prevented.

A newly recognized form of behavioral addiction is Internet gaming disorder (IGD). Experts estimate that more than 90% of U.S. children and teenagers play video games, and somewhere between 1 and 9% of them may be addicted to gaming (Gentile et al., 2017). It is important to note that—as with other addictions—heavy game play alone does not necessarily mean that an individual is addicted. Rather, individuals with IGD are preoccupied with gaming to the extent that it has become the dominant activity in daily life; they lose interest in real-life relationships and other forms of entertainment, and they may exhibit withdrawal symptoms if access to gaming is taken away.

Behavioral addictions seem to entail dysfunction in the same brain substrates that are affected by drug addictions. Functional neuroimaging (fMRI) research has documented that gambling activates the brain in much the same pattern as seen when people with cocaine addiction receive an infusion of the drug (Breiter, Aharon, Kahneman, Dale, & Shizgal, 2001). Similarly, mice that have been selectively bred for overeating show greater concentrations of dopamine in the dorsal striatum, as do mice that have been selectively bred for excessive exercising (Mathes et al., 2010).

Other evidence for the similarity between drug addiction and behavioral addiction comes from patients with Parkinson’s disease, who gradually lose dopamine-producing neurons. Patients with Parkinson’s disease often benefit from treatment with dopamine agonists, drugs that “mimic” dopamine by activating dopamine receptors in the brain. Curiously, a small percentage of patients treated with high doses of dopamine agonists suddenly develop serious gambling problems (Santangelo, Barone, Trojano, & Vitale, 2013). Apparently, in these individuals, the dopamine agonists disrupt the brain’s “wanting” system in much the same way as shown in Figure 5.15 for amphetamine and cocaine, reinforcing the problem behavior. Often, the gambling problem can be cured simply by switching the patient to a different type of medication.

Together, these results suggest there is a general reinforcement system in the brain that is activated in similar ways for different categories of reinforcers, including primary reinforcers (food and water), secondary reinforcers (money), drugs (cocaine and amphetamine), and behavior (gambling and gaming) (Breiter et al., 2001). If so, a better understanding of the biochemical and behavioral principles underlying drug addiction may also help in the treatment of individuals suffering from behavioral addictions.

**Treatments for Addiction**

The intensity of the cravings experienced in addiction, not to mention the often excruciating physical pain of withdrawal, make it impossible for many people to break free of addiction without outside help. Currently, in the United States, the vast majority of
treatment plans for addiction involve cognitive therapy, which can include one-on-one sessions with a psychologist as well as self-help sessions with a support group (such as Alcoholics Anonymous and its many spin-off organizations—Gamblers Anonymous, Narcotics Anonymous, Overeaters Anonymous, and so on). Medical treatment may also help. For example, naltrexone is a drug that blocks opiate receptors, presumably decreasing the ability of heroin to bind to those receptors and cause a hedonic reaction. In some studies, people with heroin addiction who had undergone detoxification were able to stay off heroin longer if they received continuing treatment that included naltrexone (Kirchmayer et al., 2002; Rawson & Tennant, 1984). Compulsive gamblers have also reported reductions in gambling urges following naltrexone treatment (Grant, Kim, & Hartman, 2008).

One way to interpret the effect of naltrexone is by considering addiction as a strong $S^D \rightarrow R \rightarrow O$ association in which sets of environmental stimuli ($S^D$) trigger the addictive behavior ($R$), resulting in the reinforcing outcome ($O$) of a “high” or a reduced craving or both. If the response $R$ no longer produces the outcome $O$, then the frequency of $R$ should decline. Naltrexone works by blocking the brain’s reinforcement system, so that subsequent heroin (or gambling) will no longer produce the outcome that it used to. As a result, the response (drug taking or gambling) should extinguish.

Unfortunately, treatment with naltrexone has drawbacks as well. One is that naltrexone must be taken daily. Another is that it does not reduce the person’s craving (“wanting”) for heroin. As a result, a recent review of treatments for people addicted to heroin or other opiate drugs concluded that oral naltrexone treatment was not statistically superior to placebo in preventing relapse (Minozzi et al., 2011).

The principles of operant conditioning have also suggested several other behavioral therapies to help fight addiction. One conditioning-based technique for combating addictions and habits is distancing, or avoiding the stimuli that trigger the unwanted response. For example, a cigarette smoker who is struggling to quit and who gets the urge to light up whenever she hangs out with friends who are heavy smokers should try to avoid those situations. If the $S^D$ is never present, the $R$ may never be triggered.

Another technique is differential reinforcement of alternative behaviors (DRA), which you read about earlier in this chapter. If a smoker makes it through a whole week without a cigarette, she can reinforce her own abstinent behavior by treating herself to a favorite food or activity. Friends can help by praising the nonsmoking behavior; one aspect of Alcoholics Anonymous is the social reinforcement provided by the group for each week the alcoholic stays sober. Some programs for heroin addicts reinforce abstinence with actual monetary vouchers, providing yet another form of reinforcement for the alternative behavior of not using the drug (Preston, Umbricht, & Epstein, 2000).

A final conditioning-inspired technique is delayed reinforcement: whenever the smoker gets the urge to light up, she can impose a fixed delay (e.g., an hour) before giving in to it. Recall that increasing the delay between response and outcome weakens learning (Figure 5.4). Imposing long delays between cravings and cigarettes may similarly weaken the association; it will also, by default, reduce the total number of cigarettes smoked per day.

These and other approaches can be used in combination to increase the chances of success. But even with all these therapies, addicted cigarette smokers...
(and alcoholics and drug users and gamblers) can still have a very hard time kicking their habits. Currently, the most successful approaches appear to be those that combine cognitive therapy (including counseling and support groups) with behavioral therapy based on conditioning principles—along with medication for the most extreme cases (Grant et al., 2010).

**Interim Summary**

- Addictive drugs hijack the brain’s reinforcement system and can produce psychological as well as physiological addiction.
- Opiates such as heroin and morphine mimic the endogenous opioids that are believed to form the brain’s “liking” system, while drugs such as amphetamine and cocaine affect the dopamine system, which is thought to signal “wanting” or craving.
- In addition, addiction is driven by the desire to avoid the aversive symptoms of withdrawal; these aversive symptoms may be signaled by the insula.
- Behavioral addictions may reflect the same brain processes as drug addictions.
- Among the treatments for people with addictions are cognitive therapies, medications, and behavioral therapies, including those based on principles learned from operant conditioning.

**Synthesis**

Recall the mine-sniffing rats you met in the introduction to this chapter. Their trainers teach them to scamper across a minefield to detect and signal buried explosives, in return for clicker reinforcement. It is a job the small animals can perform with little risk to themselves, while potentially saving thousands of humans from injury and death. The trainers use basic principles of operant conditioning, shaping one element of the behavior at a time and reinforcing desired responses with a clicker that acts as a secondary reinforcer. The principles of operant conditioning also apply to Becky, the little girl who was reinforced by an allowance for the behavior of keeping her room clean and who might be punished by time-outs at school for aggressive behavior during recess. Becky’s parents will turn to the same principles again when it is time to housebreak a new puppy.

Operant conditioning is a powerful form of learning that can be applied to adult humans as well. You can even use it on yourself if you want to break a bad habit or reinforce good study practices. It is an approach that has been implicitly appreciated at least since the time of the ancient Greeks. In Aristophanes’s play *Lysistrata,* first performed in 411 BC, the women of Athens agree to withhold sexual favors from their men until the men call off a frivolous war with Sparta—a clear example of negative punishment.

Many brain areas participate in operant conditioning. \( S^D \rightarrow R \) associations may be stored in connections between the sensory cortex and the motor cortex and in the dorsal striatum; predictions about the outcome \( O \) that will follow \( R \) depend on frontal areas, including the orbitofrontal cortex. The brain also seems to have
a general-purpose reinforcement system that includes the dopaminergic system, which may help an organism choose between competing responses with different predicted outcomes, and the endogenous opioids, which may signal hedonic value (“yum!”). Meanwhile, the insula helps signal the punishment value of stimuli (“ugh!”) and may provide this information to the striatum to help us learn which actions will result in avoiding or mitigating that punishment.

Many kinds of addiction, including drug addictions and behavioral addictions, result when chemicals or behaviors interfere with the brain’s reinforcement system; many types of behavioral therapy for addiction apply operant conditioning procedures to help improve people’s lives. Operant conditioning also forms a basis for understanding behavioral economics, the study of how individuals allocate their time and energy among different available responses.

The bottom line: operant conditioning is not just for training rats in Skinner boxes. People use many operant conditioning techniques in daily life without even realizing it. By understanding the underlying principles, you can use them much more effectively.

Know Your Key Terms

altruism, p. 202
basal ganglia, p. 205
behavioral addiction, p. 218
behavioral economics, p. 199
bliss point, p. 199
chaining, p. 183
concurrent reinforcement schedule, p. 198
continuous reinforcement schedule, p. 194
cumulative recorder, p. 180
delay discounting, p. 201
differential reinforcement of alternative behaviors (DRA), p. 189
discrete-trials paradigm, p. 179
discriminative stimulus (S^D), p. 177
dorsal anterior cingulate cortex (dACC), p. 215
dorsal striatum, p. 205
drive reduction theory, p. 185
endogenous opioid, p. 212
fixed-interval (FI) schedule, p. 195
fixed-ratio (FR) schedule, p. 194
free-operant paradigm, p. 179
hedonic value, p. 209
incentive salience hypothesis, p. 211
insular cortex (insula), p. 214
matching law of choice behavior, p. 199
motivational value, p. 209
negative contrast, p. 185
negative punishment, p. 192
negative reinforcement, p. 191
operant conditioning, p. 175
orbitofrontal cortex, p. 206
partial reinforcement schedule, p. 194
pathological addiction, p. 216
positive punishment, p. 191
positive reinforcement, p. 191
post-reinforcement pause, p. 194
primary reinforcer, p. 185
punisher, p. 183
punishment, p. 184
reciprocal altruism, p. 203
reinforcement, p. 184
reinforcement schedule, p. 194
reinforcer, p. 183
secondary reinforcer, p. 185
self-control, p. 201
shaping, p. 182
Skinner box, p. 179
substantia nigra pars compacta (SNc), p. 210
token economy, p. 186
variable-interval (VI) schedule, p. 196
variable-ratio (VR) schedule, p. 196
ventral tegmental area (VTA), p. 209
Quiz Yourself

1. In operant conditioning, ____________ signal whether a particular response will lead to a particular outcome. (p. 177)

2. ____________ refers to actions or behaviors that benefit another individual at the expense of some cost to the actor. Some such actions may reflect ____________, or “tit-for-tat” behavior, in which one friend helps another with the expectation that the friend will reciprocate later on. (pp. 202–203)

3. The ____________ is a part of the brain that helps determine subjective values of punishers, such as whether the intense heat of a chili pepper on the tongue is perceived as pleasurable or painful. The ____________ is a part of the brain that helps determine the motivational value of punishment—that is, what we do about it. (pp. 214–215)

4. ____________ refers to a strong habit that is maintained despite harmful consequences; if the habit is a behavior, it is called a(n) ____________. (pp. 216, 218)

5. The part of the brain called the ____________ contains dopamine-producing neurons that project to the ____________, which is important for stimulus–response learning. A different area called the ____________ contains dopamine-producing neurons that project to the frontal cortex and other brain areas. (pp. 209–210)

6. The ____________ theory states that learning is driven by an organism’s biological need to reduce innate drives to obtain primary reinforcers. (p. 185)

7. In operant conditioning procedures, ____________ is the process of providing outcomes for a behavior that increase the probability of that behavior occurring again in the future, whereas ____________ is the process of providing outcomes that decrease the probability. (p. 184)

8. ____________ is the study of how organisms allocate their time and resources among possible options. (p. 199)

9. In the operant conditioning technique known as ____________, successive approximations to a desired response are reinforced. In the operant conditioning technique known as ____________, organisms are gradually trained to execute complicated sequences of discrete responses. (pp. 182–183)

10. ____________ are stimuli such as food and sleep that can function as reinforcers due to their innate biological value to the organism; if these stimuli are paired with other stimuli that have no biological value, such as money or clicking sounds, those other stimuli can become ____________. (p. 185)

11. In a(n) ____________ schedule, every instance of the response is followed by the consequence; in a(n) ____________ schedule, only some responses are reinforced. (p. 194)

12. An area of the prefrontal cortex called the ____________ is important for learning to predict which outcomes follow particular responses. (p. 206)

13. If a rat in a Skinner box expects that making a response will result in delivery of sugary water but the rat instead receives only plain water, the phenomenon of ____________ predicts that the rat will respond less than if it had received plain water for its efforts all along. (p. 185)

14. ____________ are naturally occurring neurotransmitter-like substances that may help signal hedonic value (“liking”) in the brain. (p. 212)

15. In a(n) ____________ reinforcement schedule, an organism has a choice between multiple possible responses that may each lead to different outcomes. The ____________ predicts that the organism will make each response at a rate proportional to how often that response is reinforced relative to the other choices. (pp. 198–199)

16. In ____________, organisms learn to make responses in order to obtain or avoid certain outcomes. (pp. 175–176)

17. Training paradigms that can cause responses to become less frequent over time include ____________, in which an undesirable element is delivered after a response, and ____________, in which a desirable element is taken away after the response. (pp. 191–192)

18. Training paradigms that can cause responses to become more frequent over time include ____________, in which a desirable element is delivered after a response, and ____________, in which an undesirable element is taken away after the response. (p. 191)
19. In a fixed-ratio (FR) schedule of reinforcement, organisms typically give bursts of responding leading up to each reinforcement, followed by a(n) ______ before the next bout of responding begins. (p. 194)

20. ______ refers to the subjective “goodness” of a stimulus. The amount of work an organism will be willing to do to obtain that stimulus depends on the ______ of that stimulus. (p. 209)

Concept Check

1. A new kindergarten teacher wants to train her pupils to put away toys after playtime. Suggest at least three conditioning techniques she could use to train this behavior.

2. An employer wants to start testing her employees for drugs. She hopes that the threat of such tests will encourage employees to avoid drug use. According to the principles of conditioning, what would be the best way to schedule these drug tests?

3. Imagine that the police raid a house party and find the crowd taking an unfamiliar kind of drug. The users are sitting in front of a TV, enthusiastically munching on stale bread and laughing hysterically at an old sitcom. Even without submitting samples of the drug to a laboratory, what might we hypothesize about the mechanisms by which this drug works?

4. After the 9/11 terrorist attacks on the World Trade Center in New York City, sniffer dogs were used to search for survivors among the rubble. As the days went by and no survivors were found, handlers reported that the dogs showed signs of “depression,” or “not wanting to go to work in the morning.” To combat this, at the end of a long day of failure, handlers would arrange for the dogs to find a “survivor” (really, a healthy human hiding in the rubble) so that the dogs would end the day on a high note. Without invoking cognitive or emotional concepts, use the principles of operant conditioning to explain the utility of these “arranged finds.”

Answers to Test Your Knowledge

Is It Classical or Operant?

1. Whistling is the discriminative stimulus (S^D), birds arriving is the learned response (R), and birds eating the crumbs is the outcome (O). The birds do not get the crumbs (O) unless S^D is present and they make response R, so this is operant conditioning.

2. Thunder always follows lightning. This occurs whether or not Snoopy hides. Therefore, this is classical conditioning (CS = lightning, US = thunder, CR = hiding).

3. Presence of clouds is the discriminative stimulus (S^D), bringing the umbrella is the learned response (R), and staying dry is the outcome (O). The outcome (O) does not occur unless S^D was present and Michael brought his umbrella (R), so this is operant conditioning.

Reinforcement Versus Punishment

1. (a) The learner is Lucy; (b) response R is a tantrum; (c) outcome O is obtaining candy; (d) S^D is seeing candy at the store because tantrums at other times will not cause Mother to purchase candy; (e) candy is added to Lucy’s environment (addition = positive); (f) the frequency of Lucy’s tantrum-throwing response increases—so this is a reinforcement paradigm. Conclusion: Lucy learns to throw tantrums to obtain candy. This is positive reinforcement.
2. (a) The learner is Susan; (b) response R is giving candy; (c) outcome O is stopping the tantrum; (d) SD is the child’s tantrum when she is refused candy because candy won’t stop a tantrum unless a tantrum is already present; (e) the tantrum is taken away from Susan’s environment (subtraction = negative); (f) the frequency of Susan’s candy-giving response increases—so this is a reinforcement paradigm. Conclusion: Susan learns to give candy to stop the tantrums. This is negative reinforcement.

3. (a) The learner is Snoopy; (b) response R is crossing the boundary; (c) outcome O is the noise; (d) SD is wearing the collar while in the yard because straying at other times will not produce the noise; (e) the noise is added to Snoopy’s environment (addition = positive); (f) the frequency of Snoopy’s straying response decreases—so this is a punishment paradigm. Conclusion: Snoopy learns to stop straying beyond the boundary. This is positive punishment.

4. (a) The learner is Miguel; (b) response R is drinking alcohol; (c) outcome O is revocation of playing privileges; (d) SD is football season because Miguel will not be punished by the coach at other times of the year; (e) playing privileges are taken away from Miguel (subtraction = negative); (f) the frequency of Miguel’s alcohol drinking behavior decreases—so this is a punishment paradigm. Conclusion: Miguel learns not to drink alcohol during football season. This is negative punishment.

5. (a) The learner is Rachel; (b) response R is getting a stomachache (and telling the nurse about it); (c) outcome O is avoiding gym class; (d) SD is school lunch on days when Rachel has gym class because sickness at other times does not affect gym class; (e) gym class is taken away from Rachel’s environment (subtraction = negative); (f) the frequency of Rachel’s sickness reporting behavior increases—so this is a reinforcement paradigm. Conclusion: Rachel learns to feel (or report) sickness in order to avoid gym class. This is negative reinforcement.

Reinforcement Schedules
1. FR; 5 homeworks = 1 toy. Note that the stars themselves are actually secondary reinforcers (on an FR 1 schedule, since 1 homework = 1 star).
2. VR; 20 phone calls = 2 sales on average, but the telemarketer can’t be sure exactly when the next sale will come; it might come immediately after a previous call that also resulted in a sale.
3. VI; the couple can’t be sure exactly when their table will be ready, so they check back periodically; only the last response after the table becomes ready will be reinforced. We would expect the couple to check infrequently in the first 20 minutes and more frequently as the expected 30-minute interval nears its end.
4. FI; money can be obtained for the first donation after a 2-week interval.
5. VI; the surfer can’t be sure exactly when the next big wave will arrive, so the best policy is to hang around in anticipation, even immediately after a previous big wave.
6. FR; 1 gum = 1 negative reinforcement, which is removal of bad breath. This is also an example of continuous reinforcement, since each response R elicits the outcome O.
7. FI; the quiz occurs after a fixed 1-week interval, and good grades reinforce the studying behavior.
8. VR; 100 cards = 1 winner on average, but the very next card after a winner might also be a winner.