7

Instrumental or Operant Conditioning

Did you know that

- learning a new instrumental response often involves putting familiar response components into new combinations?
- variability in behavior is a great advantage in learning new responses?
- the deleterious effects of reinforcement delay can be overcome by presenting a marking stimulus immediately after the instrumental response?
- Thorndike's law of effect does not involve an association between the instrumental response and the reinforcer?
- instrumental conditioning can result in the learning of four binary associations and one hierarchical association?
- Pavlovian associations acquired in instrumental conditioning procedures can disrupt performance of instrumental responses, creating biological constraints on instrumental conditioning?
- the various associations that develop in instrumental conditioning are difficult to isolate from each other, which is a challenge for studying the neurobiology of instrumental learning?

The various procedures that we described so far (habituation, sensitization, and Pavlovian conditioning) all involve presentations of different types of

https://doi.org/10.1037/0000363-007
The Essentials of Conditioning and Learning, Fifth Edition, by M. Domjan and A. R. Delamater

stimuli according to various arrangements. The procedures produce changes in behavior—increases and decreases in responding—as a result of these presentation schedules. Although they differ in significant ways, an important common feature of habituation, sensitization, and Pavlovian conditioning procedures is that they are administered independently of the actions of the organism. What the participants do as a result of the procedures does not influence the stimulus presentation schedules.

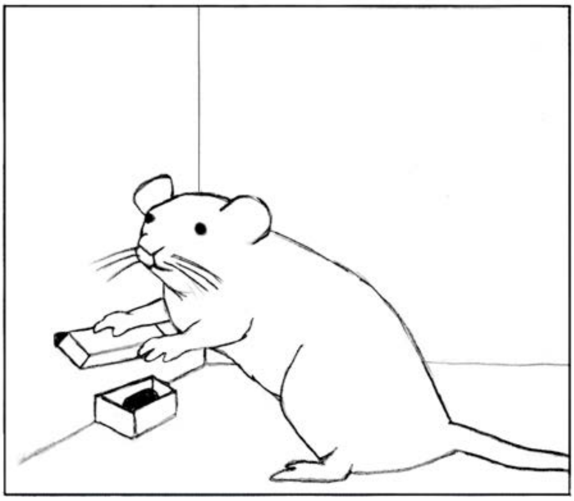
In a sense, studies of habituation, sensitization, and Pavlovian conditioning represent how organisms learn about events that are beyond their control. Adjustments to uncontrollable events are important because many aspects of the environment are beyond our control. What day a class is scheduled, how long it takes to boil an egg, how far it is between city blocks, and when the local grocery store opens are all beyond our control. Although learning about uncontrollable events is important, not all learning is of this sort. Another important category of learning involves situations in which the occurrence of a significant event or unconditioned stimulus depends on the individual's actions. Such cases involve instrumental conditioning.

In **instrumental conditioning** procedures, whether a significant stimulus or event occurs depends on the behavior of the participant. Common examples of instrumental behavior include pulling up the covers to get warm in bed, putting ingredients together to make lemonade, changing the TV channel to find a more interesting show, and saying hello to someone to get a greeting in return. In all these cases, the individual has to perform an action to obtain a specific result or outcome. Because the action is instrumental in producing the outcome, the action is referred to as **instrumental behavior**. The consequent outcome (the warmth, the tasty lemonade, the TV show, the reciprocal greeting) is referred to as the **reinforcer**. An instrumental conditioning procedure sets up a **response–reinforcer contingency**, according to which the reinforcer is delivered if and only if the target instrumental response has been performed.

Operant conditioning is a form of instrumental conditioning in which the response required for reinforcement is an operant response, identified by its effect in manipulating the environment in some way. **Operant behavior** is defined by how the behavior changes the environment. For example, turning a doorknob sufficiently to open a door is an operant response because it changes the status of the door from being closed to being open. In identifying instances of this operant response, it does not matter whether the doorknob is turned with a person's right hand, left hand, fingertips, or with a full grip of the knob. Such variations in response topography are ignored in studies of operant behavior. The focus is on the common environmental change (opening the door) that is produced by the operant behavior.

A familiar example of operant behavior in animal research involves a laboratory rat pressing a response lever in a small experimental chamber (see Figure 7.1). Whether a lever-press has occurred can be determined by placing a microswitch under the lever. In a typical experiment, presses of the lever with enough force to activate the microswitch are counted as instances of the

FIGURE 7.1. A Common Laboratory Preparation for the Study of Operant Behavior



Note. The drawing shows a rat in a lever-press chamber. A food cup is located below the lever.

lever-press operant. Whether the rat presses the lever with its right or left paw or its nose is ignored as long as the microswitch is activated. Another common example of operant behavior in animal research is a pigeon pecking a disk or a stimulus on a wall. Various ways of pecking are ignored provided that the pecks are detected by the touchscreen on the wall.

THE TRADITIONS OF THORNDIKE AND SKINNER

The intellectual traditions of classical conditioning were established by one dominant figure, Ivan Pavlov. In contrast, the intellectual traditions of instrumental or operant conditioning have their roots in the work of two American giants of 20th-century psychology, Edward L. Thorndike and B. F. Skinner. The empirical methods as well as the theoretical perspectives of these two scientists were strikingly different, but the traditions founded by each of them have endured to this day. We first consider the distinctive experimental methods used by Thorndike and Skinner and then note some differences in their theoretical perspectives.

Thorndike was interested in studying animal "intelligence." To do this, he designed several puzzle boxes for young cats in a project that became his PhD dissertation (Thorndike, 1898). A **puzzle box** is an experimental chamber used to study instrumental conditioning, in which the participant must perform a specified behavior to be released from the box and obtain a reinforcer. The puzzle for the kitten was to figure out how to escape from the box and obtain food.

Thorndike would put a kitten into a puzzle box on successive trials and measure how long the kitten took to get out of the box and obtain a piece of fish. In some puzzle boxes, the kittens had to make just one type of response to get out (e.g., turning a latch). In others, several actions were required, and these had to be performed in a particular order. Thorndike found that with repeated trials, the kittens got quicker and quicker at escaping from the box. Their escape latencies decreased.

The Discrete-Trial Method

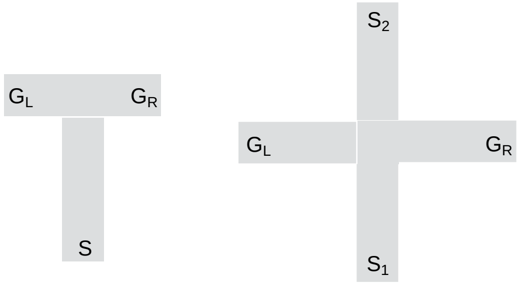
Thorndike's experiments illustrate the discrete-trial method used in the study of instrumental behavior. In the **discrete-trial method**, participants can perform the instrumental response only at certain times (during discrete trials) as determined by the experimenter. In the case of Thorndike's experiments, the kitten could only perform the instrumental escape response after it was placed in a puzzle box. When it made the required response, it was released from the box and got a piece of food. The next trial did not begin until Thorndike decided to put the kitten back in the box.

The discrete-trial method was subsequently adopted by investigators who used mazes of various sorts to study instrumental conditioning. Mazes are typically used with laboratory rats and mice. They were added to the toolbox of behavioral scientists by Willard Small, who built a maze in an effort to mimic the tunnel-like structures of the underground burrows in which rats live (Small, 1899, 1900).

Figure 7.2 shows two mazes that remain in use in contemporary research. The maze on the left is a **T-maze**, a maze with a start box that opens to a straight alley, at the end of which the participant can turn either right or left to reach the goal box. When the door to the start box is lifted, the rat can go down the stem of the maze to the choice point, where it can turn right or left to obtain a pellet of food.

The **plus maze**, shown on the right of Figure 7.2, is a variation on the T-maze, with two start boxes on opposite ends of the central alley. During the initial

FIGURE 7.2. Top View of a T-maze and a Plus Maze



Note. G = goal box; S = start box.

phase of training, each trial starts with the rat in the same start box (S_1, let's say). As in the T-maze, a left turn may be the reinforced response (with food presented in G_L). After the rat has become proficient in making the left turn from S_1, it is given a test trial that starts with the rat placed in the opposite start box, S_2. The advantage of starting the test trial from the opposite start box is that it permits determining whether the rat learned to make a particular response (turn left) or learned to go to a particular place (G_L). If the rat learned a particular response (turn left), it will turn left when started from S_2 and end up in G_R. If the rat learned to go to a particular place, it will end up in the same goal box (G_L) where it found food during the training phase. Place generally governs responding early in training, with response learning dominating late in training (Packard, 2009; Packard & McGaugh, 1996).

The discrete-trial method requires numerous manipulations. The experimenter has to pick up the rat, place it in the start box, wait for it to reach the goal box, remove it from the goal box, and put the rat in a holding area for the intertrial interval. The experimenter then has to decide how long to keep the rat in the holding area before starting the next trial.

The Free-Operant Method

The major alternative to the discrete-trial method for the study of instrumental behavior is the **free-operant method**, which permits repetitions of the instrumental response at any time. The free-operant method was developed by B. F. Skinner (1938). Skinner made numerous contributions, both methodological and conceptual, to the study of behavior, and these two types of contributions were often interrelated. The free-operant method is a case in point.

Skinner's development of the free-operant method began with an interest in designing an automated maze for rats—a maze in which the rats would automatically return to the start box after each trial. Such an apparatus would have the obvious advantage that the experimenter would have to handle the rat only at the start and the end of a training session, making it easier to conduct training sessions. An automated maze would also permit the rat rather than the experimenter to decide when to start its next trial. This would permit the investigation of not only how rapidly the rat completed an instrumental response but how frequently it elected to perform the response. Thus, an automated maze promised to provide new information about the rat's motivation that was unavailable with the discrete-trial method.

Skinner tried several approaches to automating the discrete-trial maze procedure. Each approach incorporated some improvements on the previous design, but as the work progressed, the apparatus became less and less like a maze (Skinner, 1956). The end result was what has come to be known as the Skinner box, in which a rat presses a response lever to obtain a piece of food that is dropped in a food cup nearby (Figure 7.1).

In the Skinner box, the response of interest is defined in terms of the closure of a microswitch. The computer interface ignores whether the rat presses the lever with one paw or the other, or with its tail. Another important feature of the Skinner box is that the operant response can occur at any time during a training session. The interval between successive responses is determined by the rat rather than by the experimenter. Because the operant response can be made at any time, the method is called the free-operant method.

The primary conceptual advantage of the free-operant method is that it allows the participant to initiate the instrumental response repeatedly. Skinner focused on this aspect of behavior. How often a rat initiates the operant response can be quantified in terms of the frequency of the response in a given period of time, or the **rate of responding**. Rate of responding soon became the primary measure of behavior in experiments using the free-operant method.

THE INITIAL LEARNING OF AN INSTRUMENTAL OR OPERANT RESPONSE

People often think about instrumental or operant conditioning as a technique for training new responses. Swinging a bat, throwing a football, or playing the drums all involve instrumental responses that skilled players learn through extensive practice. However, in what sense are these responses new? Does instrumental conditioning always establish entirely new responses? If not, what is the alternative? The alternative interpretation is that instrumental conditioning may involve learning to combine familiar responses in new ways or learning to make familiar responses in a new situation.

Learning Where and What to Run For

Consider, for example, a hungry rat learning to make a left turn in a T-maze to obtain a piece of food. The rat may be slow to reach the goal box at first. This is not, however, because it enters the experiment without the motor skill of getting from the start box to the goal box. Experimentally naive rats do not have to be taught to run or to turn one way or the other. What they have to be taught is where to run, where to turn, and what they will find in the goal box. In the T-maze, the instrumental conditioning procedure provides the stimulus control and the motivation for the locomotor response. It does not establish locomotion as a new response in the rat's repertoire.

Constructing New Responses From Familiar Components

The instrumental response of pressing a lever is a bit different from running. An experimentally naive rat has probably never encountered a lever before and never performed a lever-press response. Unlike running, lever pressing has to be learned in the experimental situation. But does it have to be learned from scratch? Hardly.

An untrained rat is not as naive about pressing a lever as one might think. Lever pressing consists of several components: balancing on the hind legs, raising one or both front paws, extending a paw forward over the lever, and then bringing the paw down with sufficient force to press the lever and activate the microswitch. Rats perform responses much like these at various times while exploring their cages, exploring each other, or handling pieces of bedding or pellets of food. What they have to learn in the operant conditioning situation is how to put the various response components together to create the new lever-press response.

Pressing a lever is a new response only in the sense that it involves a new combination of response components that already exist in the participant's repertoire. In this case, instrumental conditioning involves the construction or synthesis of a new behavioral unit from preexisting response components (Balsam et al., 1998).

Shaping New Responses

Can instrumental conditioning also be used to condition entirely new responses, responses that an individual would never perform without instrumental conditioning? Most certainly. Instrumental conditioning is used to shape remarkable feats of performance in sports, ice skating, ballet, and music–feats that almost defy nature. A police dog can be trained to climb a 12-foot vertical barrier, a sprinter can learn to run a mile in under 4 minutes, and a golf pro can learn to drive a ball 200 yards in one stroke. Such responses are remarkable because they are unlike anything someone can do without special training.

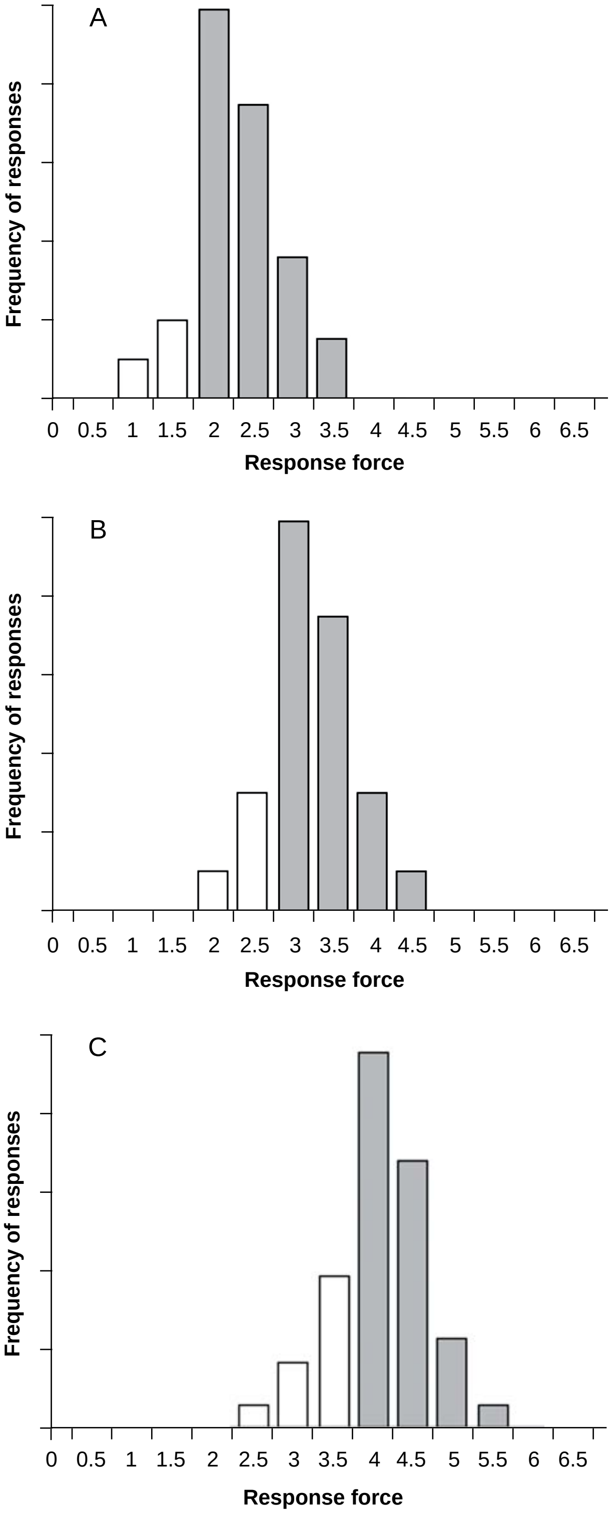
In an instrumental conditioning procedure, the individual has to perform the required response before the outcome or reinforcer is delivered. Given this restriction, how can instrumental procedures be used to condition responses that never occur on their own? The learning of entirely new responses is possible because of the variability of behavior. Variability is perhaps the most obvious feature of behavior. Organisms rarely do the same thing twice in exactly the same fashion. Response variability is usually considered a curse because it reflects lack of precision and makes predicting behavior difficult. However, for learning new responses, variability is a blessing.

With instrumental conditioning, the delivery of a reinforcer (e.g., a pellet of food) does not result in repetition of the same exact response that produced the reinforcer. If a rat, for example, is reinforced with a food pellet for pressing a lever with a force of 2 grams, it will not press the lever with exactly that force thereafter. Sometimes it will respond with less pressure, other times with more.

The panel A in Figure 7.3 shows what the distribution of responses might look like in an experiment where lever pressing is reinforced only if a force greater than 2 grams is used. Notice that many, but not all, of the responses exceed the 2-gram criterion. A few of the responses exceed a force of 3 grams, but none exceeds 4 grams.

Because the variability in behavior includes responses as forceful as 3 grams, we can change the response criterion so that the reinforcer is now only provided if the rat presses the lever with a force exceeding 3 grams. After several

FIGURE 7.3. Frequency of Lever-Press Responses Involving Various Degrees of Force



Note. In panel A, only responses greater than 2 grams in force resulted in delivery of the reinforcer. In panel B, only responses greater than 3 grams in force were reinforced. In panel C, only responses greater than 4 grams in force were reinforced. (Data are hypothetical.)

sessions with this new force requirement, the distribution of lever presses will look something like what is shown in panel B of Figure 7.3.

Responding remains variable after the shift in the response requirement. Increasing the force requirement shifts the force distribution to the right so that the majority of the lever presses now exceed 3 grams. One consequence of this shift is that the rat occasionally presses the lever with a force of 4 grams or more. Notice that these responses are entirely new. They did not occur originally.

Since we now have responses exceeding 4 grams, we can increase the response requirement again. We can change the procedure so that now the food pellet is given only for responses that have a force of at least 4 grams. This will result in a further shift of the force distribution to yet higher values, as shown in panel C of Figure 7.3. Now most of the responses exceed 4 grams, and sometimes the rat presses the lever with a force greater than 5 grams. Responses with such force never occurred at the start of training.

The procedure we just described is called **shaping**, a concept we first introduced in Chapter 2. Shaping is used when the goal is to condition instrumental responses that are not in the participant's existing repertoire. New behavior is shaped by imposing a progressive series of response requirement. The progressive response requirements gradually take the participant from its starting behavioral repertoire to the desired target response (e.g., Deich et al., 1988; Galbicka, 1988; Stokes et al., 1999).

In setting up a shaping procedure, the desired final performance must be clearly defined. This sets the goal or end point of the shaping procedure. Next, the existing behavioral repertoire of the participant has to be documented so that the starting point is well understood. Finally, a sequence of training steps has to be designed to take the participant from its starting behavior to the final target response. The sequence of training steps involves successive approximations to the final response. Therefore, shaping is typically defined as the *reinforcement of successive approximations.*

Shaping is useful not only in training entirely new responses but also in training new combinations of existing response components. Riding a bicycle, for example, involves three major response components: steering, pedaling, and maintaining balance. Children learning to ride usually start by learning to pedal. Pedaling is a new response. It is unlike anything a child is likely to have done before getting on a bicycle. To enable the child to learn to pedal without having to balance, a child starts on a tricycle or a bicycle with training wheels. While learning to pedal, the child is not likely to pay much attention to steering and will need help to make sure she does not drive into a bush or off the sidewalk.

Once the child has learned to pedal, she is ready to combine this with steering. Only after the child has learned to combine pedaling with steering is she ready to add the balance component. Adding the balance component is the hardest part of the task. That is why parents often wait until a child is proficient in riding a bicycle with training wheels before letting her ride without them.

THE IMPORTANCE OF IMMEDIATE REINFORCEMENT

Instrumental conditioning is basically a response selection process. The response (or unique combination of response components) that is required to obtain the reinforcer is selected from the diversity of actions the organism performs in the situation. It is critical to this response selection process that the reinforcer be delivered immediately after the desired or target response. Providing the reinforcer immediately after the target response in instrumental conditioning is referred to as **response–reinforcer contiguity**. If the reinforcer is delayed after the target response, other activities are bound to intervene before the reinforcer, and one of these other activities may be reinforced instead of the target response (see Figure 7.4).

Delivering a primary reinforcer immediately after the target response is not always practical. For example, the opportunity to play on a playground is an effective reinforcer for children in elementary school. However, it would be disruptive to allow a child to go outside each time they finished a math problem. A more practical approach is to give the child a star for each problem completed, and then allow these stars to be exchanged for the chance to go to the playground. With such a procedure, the primary reinforcer (access to the playground) is delayed after the instrumental response, but the instrumental response is immediately followed by a stimulus (the star) that is associated with the primary reinforcer.

A stimulus that is associated with a primary reinforcer is called a **conditioned reinforcer,** or **secondary reinforcer** (Donahoe & Palmer, 2022). Conditioned reinforcers are established through Pavlovian associations. The delivery of a conditioned reinforcer immediately after the instrumental response overcomes the ineffectiveness of delayed reinforcement in instrumental conditioning (e.g., Winter & Perkins, 1982).

The ineffectiveness of delayed reinforcement can also be overcome by presenting a marking stimulus immediately after the target response. A **marking stimulus** is a brief visual or auditory cue that distinguishes the target response from the other activities the participant is likely to perform during a delay interval. It is not a conditioned reinforcer and does not provide information about a future opportunity to obtain primary reinforcement. The marking stimulus makes

FIGURE 7.4. Diagram of Immediate and Delayed Reinforcement of the Target Response R_x

R_1 R_2 R_3 R_4 R_x O	R_1 R_2 R_x R_3 R_4 O
Immediate reinforcement	Delayed reinforcement

Note. R_1, R_2, R_3, and so on represent different activities of the organism. O represents delivery of the reinforcer. Notice that when reinforcement is delayed after R_x, other responses occur closer to the reinforcer.

the instrumental response more memorable and helps overcome the deleterious effect of the reinforcer delay (Feng et al., 2016; B. A. Williams, 1999).

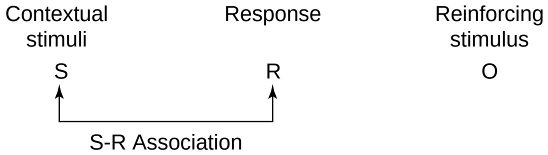
THE S–R ASSOCIATION AND THORNDIKE'S LAW OF EFFECT

Having discussed major procedural issues in instrumental conditioning, we now turn to consider underlying associative mechanisms. Instrumental conditioning is typically distinguished from Pavlovian conditioning based on procedural differences. A response is required for delivery of the reinforcer in instrumental conditioning but not in Pavlovian conditioning. The two forms of conditioning may also be distinguished based on some (but not all) of the associations that are learned in instrumental and Pavlovian conditioning.

In addition to developing the first experimental studies of instrumental conditioning, Thorndike provided a conceptual framework for analyzing instrumental conditioning that has influenced how we think about instrumental learning ever since. Although instrumental conditioning procedures emphasize the relation between the instrumental response **R** and the reinforcer or outcome **O** that follows the response, Thorndike recognized that there is a third critical component to instrumental conditioning. Instrumental responses are not performed in a vacuum. Rather, they occur in the presence of distinctive stimuli. In Thorndike's experiments these stimuli were provided by the puzzle box in which a kitten was placed at the start of a training trial. Each puzzle box had distinctive features. Once a kitten was assigned to a puzzle box, it experienced a unique set of cues whenever it performed the required escape response. Those stimuli may be represented by **S**.

Thorndike proposed that during the course of instrumental conditioning, an association comes to be established between the environmental stimuli S that are encountered when the response is made and the response R itself (see Figure 7.5). In fact, Thorndike believed that this **S–R association** was the only thing that was learned in instrumental conditioning. He summarized his thinking in the **law of effect**, which states that when a response is followed by a "pleasant state of affairs," an association is formed between the stimuli S in the presence of which the response is performed and the instrumental response R. The reinforcer delivered after the response serves to strengthen or "stamp in" this S–R association.

FIGURE 7.5. Diagram of the S–R Association in Instrumental Conditioning



Note. O = outcome; R = response; S = stimuli.

Thorndike's law of effect is counterintuitive and often incorrectly characterized. Note that according to the law of effect, instrumental conditioning does not involve learning to associate the response with the reinforcer or reinforcing outcome. Rather, instrumental conditioning results only in the establishment of an S–R association. The reinforcing outcome O is significant as a catalyst for the learning of the S–R association but is not a part of that association (e.g., Colwill & Rescorla, 1985).

The law of effect is an adaptation of the concept of elicited behavior to instrumental learning. Elicited behavior is a response to a particular stimulus. In an analogous fashion, the law of effect considers the instrumental response R to be a response to the stimulus context S in which the response is performed. The law of effect thus provided a fairly straightforward causal account of instrumental behavior.

Although it was proposed more than a century ago, the law of effect remains prominent in contemporary analyses of behavior. For example, it is used in the prominent neurocomputational discipline known as Reinforcement Learning (Sutton & Barto, 2018). Because S comes to produce R without any intervening processes, the S–R association of the law of effect has come to be regarded as the primary mechanism responsible for the habitual actions that people perform automatically without much thought or deliberation, such as brushing teeth or putting on your slippers when you get out of bed. There is considerable contemporary interest in the nature of habits and how they are learned, both among scientists and the general public. Much of the contemporary analysis of habit learning has its intellectual roots in Thorndike's law of effect (Duhigg, 2012; Perez & Dickinson, 2020; Wood et al., 2022).

Even though scientists were familiar with the law of effect throughout the 20th century, nearly 80 years passed before they started to identify the circumstances under which instrumental conditioning results in the learning of an S–R association. As we noted earlier, a critical aspect of the law of effect is that the reinforcing outcome is not a part of the S–R association that is learned. The response occurs just because the participant encounters S, not because the participant thinks about the outcome O when making the response. Thus, according to the law of effect, instrumental behavior is not "goal-directed."

How can we demonstrate that instrumental behavior is not goal-directed? If the instrumental response is not performed because of the anticipated goal or outcome, then changing how much one likes the goal should not change the rate of the instrumental response. That prediction is the basis for contemporary studies of habit learning. The experimental design is outlined in Figure 7.6. In the initial phase of the experiment, an instrumental response is established in the standard fashion. For example, a rat may be conditioned to press a response lever for food pellets flavored with sucrose. Rats like sucrose pellets and will rapidly learn to make a response (R) that produces such pellets (O).

In the second phase of the procedure, the rats are conditioned to dislike the sucrose pellets using a taste aversion conditioning procedure. They are permitted to eat the sweet pellets but are then made sick to condition the

FIGURE 7.6. Diagram of Reinforcer Devaluation Test of Goal-Directedness of Instrumental Behavior

	Initial Training	Devaluation	Test
Experimental Group	Lever Press → Sucrose Pellet	Sucrose Pellet → Illness	Lever Press
Control Group	Lever Press → Sucrose Pellet	Illness/Sucrose Pellet Unpaired	Lever Press

Note. Two groups of participants are conditioned to press a response lever for food. In the next phase, a taste aversion is conditioned to the food for the Experimental Group but not the Control Group. Both groups are then tested for lever pressing in the last phase of the experiment.

aversion. This manipulation is called **reinforcer devaluation** because it reduces the value of the reinforcer. During the reinforcer devaluation phase, the response lever is not available. Thus, reinforcer devaluation just changes the value of the reinforcer and does not involve learning anything new about the response R. For a control group, the sucrose pellets are presented unpaired with illness so that a taste aversion to the pellets is not established.

The effects of reducing the value of the reinforcer are tested in the last phase of the experiment, where the rats are again allowed to press the response lever. Here the lever pressing occurs without the sucrose pellets so that we can see whether the response is motivated by thinking about the goal object. If lever-pressing is mediated by an S–R association independent of the goal object, then reducing the value of the goal should not change the rate of pressing the bar. This follows from the fact that O is not part of what gets learned in an S–R association. However, if lever-pressing is goal directed, then less responding should occur if the value of the goal object has been reduced. (For more rigorous experimental designs to test the effects of reinforcer devaluation, see Colwill & Rescorla, 1985; Kosaki & Dickinson, 2010.)

Initial studies of the effects of reinforcer devaluation were consistent with the law of effect: instrumental responding was not reduced by reducing the value of the reinforcer. However, an important caveat was that responding became independent of the value of the reinforcer only after extensive training e.g., Dickinson, 1985). Thus, instrumental behavior is thought to be controlled by S–R associations only after extensive training. Early in training, the behavior was goal-directed and could be reduced by reinforcer devaluation.

The independence of habitual behavior from the goal of the behavior is evident in common experience. We wear shoes to protect our feet from injury when walking outside. However, we don't think about this benefit when we put on shoes before venturing outside. We brush our teeth to remove plaque and prevent cavities, but we don't think about plaque and cavities every time we use our toothbrush. Going through a cafeteria line, we automatically pick up a tray and utensils without much thought about what we might put on the tray or what utensils we might need to eat our food. If you are in the habit of getting a cup of coffee every morning, you are likely to do this without thinking about the pharmacological effects of caffeine each time.

From a historical perspective, there is a bit of irony in what we have learned about S-R associations in instrumental conditioning. On the one hand, this research has confirmed Thorndike's law of effect. However, because S–R control of instrumental behavior requires extensive training, S–R associations cannot explain the emergence of instrumental behavior early in training. Therefore, the law of effect could not have been responsible for the learning that Thorndike observed in his puzzle box experiments.

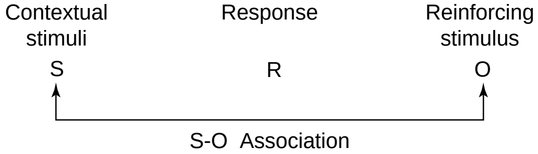
More recent research has shown that whether instrumental responding is goal-directed or mediated by an S–R habit is much more complicated than investigators initially thought. Extensive training does not always make an instrumental response habitual (Garr et al., 2021). Other relevant variables are the delay of reinforcement, predictability of reinforcement, and the density and schedule of reinforcement (e.g., Bouton, 2021; Garr et al., 2020; Urcelay & Jonkman, 2019). Whether instrumental behavior is mediated by S–R associations also depends on the individual's level of stress and anxiety (Pool et al., 2022).

ASSOCIATIVE MECHANISMS OF GOAL-DIRECTED BEHAVIOR

For behavior to be goal-directed, you have to think about the goal object when you are making the response. What might get you to think about the goal object? To answer this question, let's return to the three fundamental elements that Thorndike identified as essential to any instrumental conditioning situation. These are the response (R), the reinforcing outcome (O), and the stimuli (S), in the presence of which the response is reinforced. Given these three elements, several associations could be learned. As we have discussed, S could become associated with R. What other associations might be learned?

Because the reinforcer O occurs in the presence of S, S could become associated with O, resulting in an **S–O association** (see Figure 7.7). According to an S–O association, being in the place (S) where the instrumental response was previously reinforced could activate the memory of the reinforcer (O). Consider, for example, playing a video game that you particularly enjoy because

FIGURE 7.7. Diagram of the S–O Association in Instrumental Conditioning



Note. O = outcome; R = response; S = stimuli.

you have learned how to get a high score in the game. Here S is the app for the game, O is the high score that you can get, and R is the behavior of playing the game. Playing the game a number of times will result in an S–O association. As a result, the next time you open the app for the game (S), you will think of the high score (O) you were able to earn playing the game. S will activate the memory of O, which can then result in the motivation to perform the goal-directed behavior. This type of effect is sometimes referred to *Pavlovian-to-instrumental transfer* (or PIT, for short). Current research continues to explore the psychological and neurobiological mechanisms of how S–O associations control instrumental behavior (see Cartoni et al., 2016; Holmes et al., 2010; Rescorla & Solomon, 1967; Trapold & Overmier, 1972).

Given the three basic elements of an instrumental conditioning situation (S, R, and O), another association that can be learned is between the response R and the reinforcer O or the **R–O association**. There is a great deal of interest in the R–O association in discussions of goal-directed behavior and its relation to the S–R habit system (e.g., Balleine, 2019; Perez & Dickinson, 2020). According to the R–O association, making the instrumental response activates the memory of the reinforcer or reinforcing outcome.

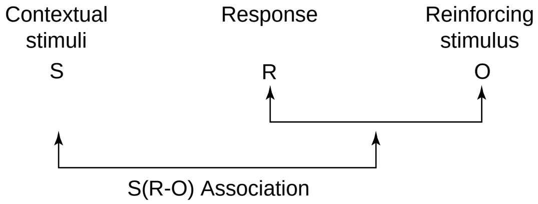
S–O and R–O associations are promising candidates for a mechanism of goal-directed behavior because they include the goal object or reinforcing outcome. However, these two associations are not sufficient by themselves to explain goal-directed behavior. The S-O association does not include the response R and therefore cannot explain why R occurs. The R-O association is more promising because it includes R. However, according to the R-O association, performing the response activates the memory of the goal object. This means that thinking about the goal does not happen until the response has occurred. That makes it difficult to explain why the response was made in the first place.

One solution to this problem was initially suggested by Pavlov (1932) and is sometimes referred to as the *bidirectional hypothesis*. In this case, Pavlov assumed that R–O associations could be formed during instrumental conditioning (e.g., when a dog learns to raise its paw for food reward). However, he also assumed that such associations could be bidirectional; thinking of O would activate R through this bidirectional association. The idea of O–R associations learned during instrumental conditioning was developed further by Trapold and Overmier (1972).

A more extensively explored solution to the problem of why the animal makes response R is provided by the hierarchical **S–(R–O) association** (see Figure 7.8). According to the S–(R–O) association, being in the place where the instrumental response was previously reinforced (S) activates the knowledge that making response R will yield the reinforcer O. The S–(R–O) associative structure attributes the occurrence of the instrumental response to S, which then activates the memory that making R will produce the reinforcer outcome.

One of the first scientists to emphasize that instrumental conditioning cannot rest on just binary relations was B. F. Skinner (1969). Skinner was not comfortable with the concept of an association or the idea that stimuli can

FIGURE 7.8. Diagram of the S(R–O) Association in Instrumental Conditioning



Note. O = outcome; R = response; S = stimuli.

activate memories. Obviously, he understood that in instrumental conditioning the presentation of reinforcer O depends on the prior occurrence of the response R. However, he also recognized that this R–O relation, is in effect only in the presence of S. Therefore, he suggested, that a relation that includes three components becomes established in instrumental conditioning: S signals that responding will produce the reinforcer or sets the occasion for the R–O relation. Skinner referred to this as a *three-term contingency*.

Experimental investigations of the associative structure of instrumental conditioning have provided evidence for a variety of associations (Rescorla, 1991): S–R, S–O, R–O, O–R, and S–(R–O). Thus, instrumental behavior is not a "simple" form of learning but involves several relationships, all of which contribute to the control of instrumental responding in different ways. The S–R relation does not include the reinforcer and therefore is not degraded by reinforcer devaluation. The reinforcer is included in each of the other relations. Therefore, those relations are all susceptible to reinforcer devaluation and may all have a role in goal-directed behavior. The S–O relation activates the memory of the reinforcer and is thus responsible for reward incentive effects. The R–O relation indicates that responding produces the reinforcer, and the S–(R–O) relation shows that this memory is activated when the individual encounters S.

IMPLICATIONS FOR BIOLOGICAL CONSTRAINTS ON INSTRUMENTAL CONDITIONING

Understanding the associative structure of instrumental conditioning helps to solve some enduring problems in instrumental learning. In some of Thorndike's puzzle boxes, the kittens had to yawn or scratch themselves to be let out (Thorndike, 1911). Learning proceeded slowly in these boxes. Even after extensive training, the kittens did not make vigorous and bona fide yawning responses; rather, they performed rapid, abortive yawns. Thorndike obtained similar results when the kittens were required to scratch themselves to be let out of the box. In this case, the kittens made rapid, halfhearted attempts to scratch themselves. These examples illustrate the general finding that self-care and grooming responses are difficult to condition with food reinforcement (Shettleworth, 1975).

Another category of instrumental behavior that is difficult to condition with food reinforcement is the release of a coin or token. Two of Skinner's graduate students, Keller and Marion Breland, became fascinated with the possibilities of animal training and set up a business that supplied trained animals for viewing in amusement parks, department store windows, and zoos. As a part of their business, the Brelands trained numerous species of animals to do various entertaining things (Breland & Breland, 1961).

For one display, they tried to get a pig to pick up a coin and drop it into a piggy bank to obtain food. Although the pig did what it was supposed to a few times, as training progressed, it became reluctant to release the coin and rooted it along the ground instead. This rooting behavior came to predominate, and the project had to be abandoned. The Brelands referred to this as "misbehavior" because it was contrary to what should have occurred based on instrumental conditioning principles. Others subsequently referred to examples of such behavior as **biological constraints on learning**.

Several factors are probably responsible for the constraints on learning that have been encountered in conditioning grooming and coin-release behavior (Shettleworth, 1975). However, one of the most important factors seems to be the development of S–O associations in these instrumental conditioning procedures (Timberlake et al., 1982). In the coin-release task, the coin becomes associated with the food reinforcer and serves as stimulus S in the S–O association. In instrumental reinforcement of grooming, stimulus S is provided by the contextual cues of the conditioning situation. Because S–O associations are much like Pavlovian associations between a conditioned stimulus and an unconditioned stimulus (US), Pavlovian conditioned responses related to the reinforcer come to be elicited by stimulus S. Pavlovian responses conditioned with food consist of approaching and manipulating the conditioned stimulus. These food-anticipatory responses are incompatible with self-care and grooming. They are also incompatible with releasing and thereby withdrawing from a coin that has come to signal the availability of food. What appears as "misbehavior" reflects the interfering effects of Pavlovian conditioned responses on instrumental behavior.

IMPLICATIONS FOR NEURAL MECHANISMS OF INSTRUMENTAL CONDITIONING

The complexity of the associative structures of instrumental learning presents serious challenges for scientists trying to discover the neural mechanisms and the neural circuitry underlying instrumental behavior. This situation seems more complex than it is for Pavlovian conditioning. As we saw in Chapters 4 and 5, there are both simple and more complex forms of Pavlovian conditioning. Simple forms of Pavlovian excitatory conditioning are mediated by S–S and S–R associations. More complex forms involve hierarchical relations, for example, as in studies of "positive occasion setting" where stimulus A is paired with the US only if it occurs following stimulus B, or B–(A–US).

However, one can examine the neural mechanisms of S–S associations with procedures that do not involve B–(A–US) relations. Unfortunately, such procedural simplification is not as easily accomplished in studies of instrumental conditioning.

Instrumental learning involves binary associations (S–R, S–O, R–O, and O–R), as well as hierarchical S–(R–O) relations. One cannot easily design an instrumental conditioning procedure that involves one of these associations to the exclusion of the others. For example, one cannot design an instrumental procedure that permits S–O associations without also allowing R–O associations, because the delivery of O contingent on R is an inherent feature of instrumental conditioning. One also cannot create an instrumental procedure that allows S–R associations without also allowing R–O and S–(R–O) associations. Finally, one cannot design a procedure that only results in an R–O association because as soon as one defines a response, one has also defined a set of cues S that occur when the response is made. Rather, the best one can do is design an experiment in which one or the other of these associations is targeted with the others controlled for.

The analysis of the neurobiology of instrumental conditioning requires deciding which associative mechanism to focus on and how to design an experiment that will isolate that particular mechanism. Success depends on sophisticated knowledge of behavior theory, in addition to expertise in neuroscience and neurobiology. This makes the study of the neural mechanisms of instrumental behavior rather difficult, but investigators are stepping up to meet the challenge (Balleine, 2019).

SUMMARY

In instrumental conditioning, the delivery of a biologically significant event or reinforcer depends on the prior occurrence of a specified instrumental or operant response. The instrumental behavior may be a preexisting response that the organism has to perform in a new situation, a set of familiar response components that the organism has to put together in an unfamiliar combination, or an activity that is entirely novel to the organism. Successful learning in each case requires delivering the reinforcer immediately after the instrumental response or providing a conditioned reinforcer or marking stimulus immediately after the response.

Instrumental conditioning was first examined in this country by Thorndike, who developed discrete-trial procedures that enabled him to measure how the latency of an instrumental response changes with successive training trials. Skinner's efforts to automate a discrete-trial procedure led him to develop the free-operant method, which allows measurement of the probability or rate of an instrumental behavior. Both discrete-trial and free-operant procedures consist of three components: contextual stimuli S, the instrumental response R, and the reinforcing outcome O. Reinforcement of R in the presence of S allows for the establishment of several binary associations (S–R, S–O, R–O, and O–R), as well as the hierarchical S–(R–O) association. These various associations are difficult to isolate from one another, which creates challenges for investigating the neurophysiology and neurobiology of instrumental learning. Moreover, the S–O association can create serious response constraints on instrumental conditioning.

SUGGESTED READINGS

Balleine, B. W. (2019). The meaning of behavior: Discriminating reflex and volition in the brain. *Neuron*, *104*(1), 47–62. https://doi.org/10.1016/j.neuron.2019.09.024

Donahoe, J. W., & Palmer, D. C. (2022). Acquired reinforcement: Implications for autism. *American Psychologist*, *77*(3), 439–452. https://doi.org/10.1037/amp0000970

Perez, O. D., & Dickinson, A. (2020). A theory of actions and habits: The interaction of rate correlation and contiguity systems in free-operant behavior. *Psychological Review*, *127*(6), 945–971. https://doi.org/10.1037/rev0000201

Wood, W., Mazar, A., & Neal, D. T. (2022). Habits and goals in human behavior: Separate but interactive systems. *Perspectives in Psychological Science*, *17*(2), 590–605. https://doi.org/10.1177/1745691621994226

TECHNICAL TERMS

instrumental conditioning, page 102
instrumental behavior, page 102
reinforcer, page 102
response–reinforcer contingency, page 102
operant conditioning, page 102
operant behavior, page 102
puzzle box, page 103
discrete-trial method, page 104
T-maze, page 104
plus maze, page 104
free-operant method, page 105
rate of responding, page 106
shaping, page 109
response–reinforcer contiguity, page 110
conditioned reinforcer, page 110
secondary reinforcer, page 110
marking stimulus, page 110
S–R association, page 111
law of effect, page 111
reinforcer devaluation, page 113
S–O association, page 114
R–O association, page 115
S–(R–O) association, page 115
biological constraints on learning, page 117

For chapter summaries and practice quizzes, visit https://www.apa.org/pubs/books/essentials-conditioning-learning-fifth-edition (see the Student Resources tab).