5

REWARD AND PUNISHMENT

All functional explanations of behavior depend on some notion of what is good and bad. If we are talking in terms of evolutionary adaptation, good and bad boil down to values of inclusive Darwinian fitness, a measure reasonably clear in principle, but often elusive in practice. If we are talking in terms of the operant behavior of individual animals, good and bad correspond to reward and punishment, to situations better or worse than the current situation. This chapter is about reward and punishment: about the concept of *reinforcement* that includes them both, about how it is defined, and the procedures used to study its effects.

There are two kinds of question to ask about reward and punishment: (a) What makes situations good and bad? Can we define good and bad independently of the behavior of the animal do all good situations share common features? Or must we always see the effect of a situation on an animal before we can be sure of its hedonic value? (b) Granted that we know the hedonic properties of a situation, how does it affect behavior? What are the mechanisms, what are the rules?

Generations of philosophically minded students of human affairs have labored over questions of the first kind, which Aristotle termed the definition of "the good." The early twentieth-century philosopher G. E. Moore summed up the modem consensus in dry but exact fashion as follows: "I have maintained that very many things are good and evil in themselves, and that neither class of things possesses any other property which is both common to all its members and peculiar to them" (Moore, 1903, p. *x*). All that good things have in common is that they are good; all that bad things have in common is that they are bad.

The early behaviorists were undeterred by the philosophers' failure to find an independent yardstick for value. Deceived by the apparent simplicity of the white rat, they tried to reduce motivated behavior to a small set of "primary drives": *Hunger, thirst,* and *sex* were on everybody's list. For the rat, at least, the definition of a good thing was that it led to the reduction of one or more of these three drives. But opinions differed about what should be done when rats sometimes did things that could not be explained by one of the three. Rats in a new environment will eventually explore it, for example; given a weak electric shock for pressing a lever, they are likely to press it again rather than avoid the lever after their first press, and so on. One school proposed new drives, like "curiosity," "habitat preference," "freedom," "sleep," and "aggression." The other school, more parsimonious, held to the original trinity and proposed to solve the problem of additional motives by linking them to the basic three. For example, exploratory behavior might be explained not by a "curiosity drive," but by a past history in which exploration had led to food, water, or sexual activity.

Neither course was wholly satisfactory. Among those willing to entertain additional drives, there was no general agreement beyond the basic three. Indeed, someone acquainted with desert animals might question even *thirst* as a primary drive, since many rarely drink in nature, obtaining the water they need from their food. The fate of the fundamentalists was no better. Although some activities could be plausibly linked to primary drives, attempts to include others appeared strained; here also there was no consensus.

As I explain in Chapter 7, the modem version of the multiple-drive view is the economic concept of a *preference structure.* This idea is both more and less ambitious than the earlier view. It is more ambitious in that it proposes to accommodate not only more than one drive (or desirable thing), but also shows how competing drives are to be reconciled. Drive theory could not easily explain how an animal both hungry and thirsty should choose between food and water, for example. It is less ambitious because it leaves on one side the issue of completeness: It is content to take situations one at a time, propose a preference structure for each, and test the implications for choice — leaving for later the issue of whether there is a global preference structure from which all operant behavior can be derived.[1]

The view that there are primary and secondary motives and that all behavior can be derived from a small primary set, has no real contemporary descendant, but recent attempts to derive a universal preference structure from a limited set of motivational *characteristics* is clearly in the same tradition (see Chapter 7).

The question of what makes a situation good or bad from an animal's point of view, interesting as it is, has not led to conclusive answers. Psychologists have been driven back to Moore's conclusion that there is nothing beyond the animal's reaction that marks a situation as good or bad. This has led to the animal-defined concept of *reinforcement,* as I describe in a moment.

The question of how situations with known hedonic properties affect behavior has attracted much more attention, for two reasons. First, it suggests numerous experiments. If we have something an animal wants, like food, and we know what the animal can do in a gross sense, then we can require all sorts of things of him as a condition for giving him the food. The way in which the animal copes with the problems we set tells us something about the machinery of reinforcement. Second, and most important, the more we know about the machinery, the closer we are to answering the first question. In science, as in education, it is as well to begin with simple, extreme cases. If we understand how animals avoid and escape from electric shock, how hungry animals cope with food problems, then we will be in a better position to understand their behavior in situations where the rewards and punishments are less obvious and their behavior accordingly more subtle.

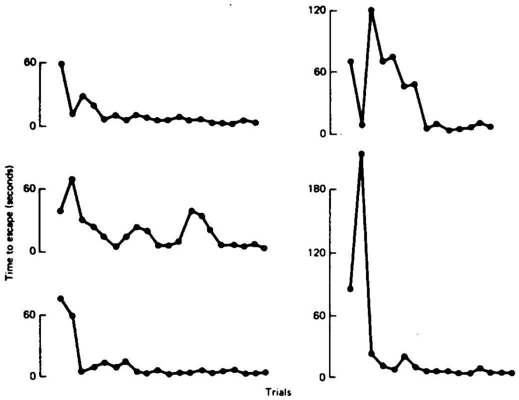


Figure 5.1. Time taken to escape from a puzzle box on successive trials by five different cats. (From Thorndike, 1898.)

This chapter discusses the main experimental arrangements that have been used to limit animals' access to food or to permit them to escape or avoid electric shock. I emphasize procedures and methods of data analysis, but also say something about how animals adapt to the procedures — later chapters expand on this. I begin with the concept of reinforcement.

REINFORCEMENT AND THE LAW OF EFFECT

The modern, experimental study of reward and punishment is usually dated from the work of Edward L. Thorndike.[2] During the last years of the nineteenth century, while a graduate student, first at Harvard University and then at Columbia, Thorndike studied the behavior of cats and other animals escaping from puzzle boxes. The cats could escape from the box by clawing on a wire loop or a bobbin, or by making some other response of this sort to unlatch the door. After each successful escape (trial) Thorndike gave the animal a brief rest, then put it in the box once again. This process was repeated until the animal mastered the task. Thorndike measured the time the animal took to escape on successive trials, producing for each a *learning curve* like the ones shown in Figure 5.1.

The learning curves in Figure 5.1 are quite variable. This is because they just measure times, and not activities. What seems to be happening is that on early trials the cat tries various things, such as pawing at the door, scratching the walls of the box, mewing, rubbing against parts of the apparatus, and so on. Most of these are ineffective in operating the latch. Because these activities occur in an unpredictable sequence from trial to trial, the effective response occurs at variable times after the beginning of a trial. Trial times improve because the ineffective acts gradually drop out.

Thorndike concentrated on trying to find the selection rule that determines how the effective act is favored over ineffective ones. He decided that *temporal contiguity* is the critical factor, together with the hedonic value of the outcome. He stated his conclusion as the well-known *law of effect:*

> Of several responses made to the same situation, those which are *accompanied or closely followed by satisfaction* to the animal. . will, other things being equal, be more firmly connected with the situation. . .; those which are accompanied or closely followed by discomfort. . will have their connections with the situation weakened.. . The greater the satisfaction or discomfort. the greater the *strengthening or weakening of the bond.* (Thorndike, 1911, p. 244, my italics)

This principle provided a framework for American studies of learning for the next sixty years. The first phrase in italics identified as critical the close temporal relation between reward (or punishment) and subsequent behavior. The term *satisfaction* identified reward and punishment as necessary for learning and raised the issue of the definition of what was subsequently to be termed *reinforcement.* The term *bond* led to the view that learning involves the formation of links or associations between particular *responses* and particular *stimuli* (situations). These three ideas have been more or less severely modified by later work. Let's look at each of them.

Obviously the law of effect would be of little use without some independent measure of what is meant by "satisfaction." If we want to train an animal according to Thorndike's law, we must know what constitutes satisfaction for it; otherwise the principle is no help. Thorndike solved this problem by making use of the fact that animals can do more than one thing to get something: A satisfying state of affairs is anything the animal "does nothing to avoid, often doing such things as to attain and preserve it." A little later, this idea was formalized as the *trans-situationality* of the law of effect: If something such as food will serve as a *reinforcer* for one response, it should also serve for others. Another way to deal with the same problem is to take *approach* and *withdrawal* as reference responses. A "satisfier" (*positive reinforcer,* in modem parlance) is something the animal will approach; a "discomforter" (*punisher, aversive stimulus,* or *negative reinforcer*[3]) is something it will withdraw from. *Reinforcers* are Pavlov's *unconditioned stimuli* (USs) discussed in the previous chapter.

Thorndike's definition of a reinforcer is the aspect of his law that has been least altered by later work. As I point out in Chapter 7, it has been extended somewhat by the notion of a preference structure, but its essential feature — that hedonic quality is revealed by the animal's own behavior — has been retained.

Comparison of Thorndike's law with the discussion of learning in the previous chapter shows that Thorndike made no distinction between local and long-term memory. Learning to escape from a puzzle box is one thing; recalling the effective response after a delay, or after being removed from the situation, is quite another. A cat may learn to escape from a puzzle box, just as *Stentor* may "learn" to escape from carmine, without being able to repeat the feat 24 hours later. The availability of a rewarding consequence is certainly necessary for adaptive behavior like this, but its role in enabling the animal to remember what it learned is not obvious. We certainly cannot assume, as Thorndike did, that reward is necessary for memory (i.e., the formation of "bonds," in his terminology). (We will see in Chapter 12 that valued events seem to be better remembered than neutral ones, but that is another matter.)

The third element in Thorndike's law is his assumption that the effective response is directly "strengthened" by its temporal contiguity with reward. It turns out that contiguity is terribly important; but it is not the only thing that is important, and Thorndike's emphasis on the single effective response at the expense of the many ineffective responses has been misleading in some ways. Michelangelo is said to have responded to a question about how he was able to sculpt so beautifully by saying: No, it is really quite easy: I just take away all the marble that is *not* the statue, and leave the rest. Thorndike's law does not sufficiently emphasize that reinforcers act by a process of selective elimination.

Later experiments have shown that response-reinforcer contiguity is not sufficient for a reinforcer to be effective, and may not always be necessary. The logic of the thing shows that strengthening-by-contiguity cannot be sufficient by itself to explain operant behavior. There are two ways to deal with this problem: One is to consider what additional processes may be necessary. The second is to look in more detail at the functional properties of operant behavior: To what procedural properties is it sensitive? In what sense do animals maximize amount of reward? The second question is much easier than the first. Moreover, the more we know about the functional properties of operant behavior, the more educated our guesses about the underlying processes can be. I follow up the question of mechanism in the last two chapters. The functional properties of reinforcement and reinforcement schedules are taken up in a preliminary way in this chapter.

Experimental methods

All science begins with taxonomy. If we want to understand the properties of reward and punishment (i.e., reinforcement), the first step is to gather some examples of how they act, and then begin classifying. But how are examples to be gathered? One could collect anecdotes: "Little Freddie used to pick his nose, but when I thrashed him soundly for doing it, he soon stopped." But this is obviously unsatisfactory: We don't know how soundly Freddie was thrashed or how soon the thrashing followed the offense or how quickly Freddie desisted. We have no precise measure of the response, the punishment, or the frequency with which one followed the other. We don't know Freddie's past history. Useful data on the effects of reward and punishment cannot be gathered like bugs at a picnic, without planning or design. We must do experiments — but what kind of experiments?

Experiments on reinforcement are of two general kinds: experiments in which the animal can improve his situation by moving about; and experiments where movement is irrelevant, but the animal can improve things by making a spatially localized response. The first category includes studies in which the animal must find food in a maze or runway or avoid electric shock in a shuttle box. Mazes are of two main sorts: the Hampton-Court variety where there is one goal box and many blind alleys, and the animal's task is to learn the one path to the goal; and the newer, radial maze, where every goal box contains food and the animal's task is to visit each goal box without repetition. Early studies of learning all used situations where locomotion was essential to reward. The second category comprises mainly so-called free-operant or Skinner-box experiments, in which the animal must press a lever or peck a lighted key for reinforcement, which is delivered in a fixed place by an automatic mechanism[4]

Maze-type experiments are useful if spatial behavior is of special interest, or if one wants to make use of animals' natural tendency to approach some things and withdraw from others. No special training is required for an animal to go from the part of a shuttle box where it has just been shocked to a part where it has never been shocked, for example. Rats will explore the goal boxes of an eight-arm radial maze without having to be led down each arm. Spatial tasks are less useful if one is interested in time relations — between reward and response, between successive responses, or between stimuli and responses. Skinner-box experiments usually require that the animal be first trained to make the instrumental response, but they are ideal for the study of time, because the experimenter can measure exactly when a particular response occurs and arrange for reward or punishment to occur in a precise temporal relation to it. The Skinner box also lends itself easily to automation: Given a food-pellet dispenser, a transducer for measuring specified aspects of the animal's behavior, and a computer to record response information, present stimuli, and operate the feeder according to a rule specified by the experimenter, human intervention is required only to place the animal in the apparatus and type "GO" on the keyboard. Since temporal relations are very important to operant behavior, and repetitive labor is irksome to most people, Skinner box experiments have become very popular.

Thorndike looked at changes in behavior across learning trials; in contemporary terms, he studied the *acquisition* of behavior. If he had persisted in running his cats even after they had mastered the task, he would have been studying *steady-state* behavior, the properties of a developed *habit.* Steady-state behavior is more interesting if reinforcement does not follow every response (*intermittent reinforcement*). It is easier to study if the animal need not be reintroduced into the apparatus after each occurrence of the reinforcer. Both these requirements favor Skinner's method over Thorndike's, and the Skinner box has become the preferred apparatus for studying steady-state operant behavior. Let's look at some common arrangements.

The Skinner box.

Skinner boxes come in many varieties. The standard version, for a rat, is a small, metal-and-Plexiglas box about 20 cm on a side. On one wall is a lever or two, often retractable, so it can be presented or withdrawn under remote control. A feeder, for either pellets or liquids, dispenses food at an aperture in the middle of the wall. Stimulus sources, in the form of a loudspeaker or buzzer, and lights above the levers, are also on the wall. The version for a pigeon is a little larger, food comes from a grain hopper, and stimuli and response transducer are combined in the form of a touch-sensitive computer screen on which colored disks (response "keys") or other visual stimuli can be presented. Pecks on the key, or presses on the lever, go to a controlling apparatus (originally a tangled mess of wires, timers, and electromagnetic relays, nowadays a personal computer) that operates the feeder and turns stimuli on or off according to the experimenter's program.

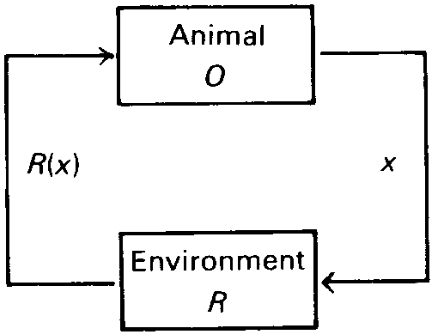


Figure 5.2. Feedback relations in an operant conditioning experiment. x = response measure; $R(x)$ = reinforcement produced by x; R = feedback function (reinforcement schedule); O = control function (behavior processes).

This basic plan can be modified in several ways. Additional transducers (for the same or different responses) can be added or the transducers might be modified for different species: Ethologists, for example, have studied Great Tits (small European perching birds) in a Skinner box with a pair of perches to record hops and a conveyor belt to present mealworms as reinforcers. In my own laboratory we have built hexagonal or octagonal enclosures for rats with transducers for different responses such as wheel running, gnawing, licking, and others in each segment.

The essential features of all these arrangements are represented in Figure 5.2, which has two parts: the programming computer, labeled as *R,* which provides reinforcers, $R(x)$, for the animal (I ignore computer-controlled stimulus changes for the moment), and the animal, labeled *0,* which provides responses, *x,* for the computer. Animal and apparatus constitute *a feedback system;* anything we measure about steady-state operant behavior, such as the animal's rate of lever pressing, or the rate at which he gets fed, will generally reflect properties of both halves of the system: the animal *and* the programming computer. Figure 5.2 is a model for all interaction between an animal and its environment (compare it with Figure 3.6). The Skinner box, which is sometimes decried by ecologically minded animal behaviorists, is nothing more than a highly controllable environment.

R and *0* are *functions: R* defines how the response the animal makes, *x,* will be translated into the reinforcers it receives *R(x). R* is of course known, since the experimenter determines the program for delivering reinforcers. Program *R* is termed a *feedback function* (or *schedule function).* Older names for *R* are *contingencies of reinforcement,* or *reinforcement schedule. 0,* the *control function,* defines how the reinforcers the animal receives will be translated into responses. Another name for *0* is the *laws, processes* or *mechanisms* of behavior. *0* is not known in general, and the aim of experiment is to help refine our understanding of it.

Figure 5.2 can be converted from an illustration to a formal model once we decide on the proper way to measure *x* (responding) and *R(x)* (reinforcer presentation: *reinforcement,* for short).[5] Chapters 6 and 7 go into this in more detail. For now let's just consider some commonly used reinforcement schedules and how the animal adapts to them.

Response- and time-based schedules of reinforcement

The simplest feedback function is when every lever press yields a food pellet. This is also the simplest *ratio schedule: fixed-ratio 1* (FR 1), also known as *continuous reinforcement.* Figure 5.3 shows how one individual, hungry (i.e., food-deprived) rat first learned to respond for food pellets delivered on a fixed-ratio 1. The rat had previously been exposed to the Skinner box and given occasional opportunities to eat from the automatic feeder, but responses to the lever had no effect. This is known as *magazine training* and just gets the animal used to eating from the feeder. On the day shown in the figure, the lever was connected to the feeder for the first time. The rat's lever presses are shown as a *cumulative record:* Time is on the horizontal axis, and each lever press increments the record on the vertical axis (each response produces only a small vertical increment, so that cumulative records appear quite smooth so long as response rate changes gradually). The first three lever presses (at time zero, close to 60 min, and about 95 min) produce food but not additional lever pressing. But at the fourth response, the animal evidently "catches on" and presses rapidly thereafter, the rapidity of his presses shown by the steepness of the record. The record begins to tip over at the extreme right of the figure, as the animal's rate of pressing slows, presumably because he is getting less and less hungry.

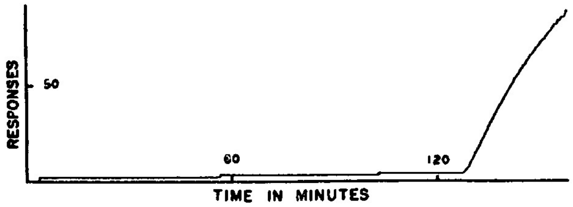


Figure 5.3. Cumulative record of the acquisition of lever pressing by a rat reinforced with food on a fixed-ratio 1 schedule. The first three feeding had little effect; the fourth is followed by a rapid increase in lever-press rate. (From Skinner, 1938, Figure 3. Reprinted by permission of Prentice-Hall, Englewood Cliffs, N.J.)

Skinner discovered that a hungry rat will continue to press a lever even if food doesn't follow every lever press. When the number of presses required for each food delivery is constant, the resulting arrangement is termed *a fixed-ratio schedule;* when the number varies from food delivery to food delivery, it is termed a *variable-ratio schedule.* The ratio value is the ratio of responses made to food deliveries received, averaged over some period of time. When the time interval involved is a single experimental session (typically 30 min to 3 hr), the relation between responses made and reinforcers received (i.e., between response and reinforcement *rates)* is known as the *molar feedback function.* For ratio schedules it takes a uniquely simple form:

$$R(x) = x/M, \qquad (5.1)$$

where *M* is the ratio value, *R(x)* the frequency of feeder presentations per unit time (food rate), and *x* the rate of lever pressing. Molar feedback functions are important for the regulation of feeding by operant behavior and for understanding the different effects of different schedules (see Chapter 7).

Fixed- and variable-ratio schedules have the same molar feedback function, but differ, of course, in their local, *molecular* properties. This difference shows up in cumulative records of steady-state (i.e., well-learned) behavior, which are shown in stylized form in Figure 5.4. The diagonal "blips" on the record indicate food (reinforcer) deliveries. The dashed horizontal lines through the FR record are separated by a constant vertical distance, to indicate that each reinforcer occurs after a fixed number of responses. Records like this have been produced by pigeons, rats, people, monkeys, and numerous other animals — performance on simple reinforcement schedules differs little across a range of mammal and bird species. Both FR and VR schedules generate high rates of responding, as shown by the steep cumulative records in the figure, but the local structure of behavior is different: Animals typically pause briefly after each food delivery on fixed-ratio schedules, but respond steadily on variable-ratio. This difference is a reaction to the fact that food never immediately follows food on FR, but sometimes does so on VR. Food *predicts* a period of no-food on FR, but if on VR the number of responses required varies randomly from one interfood interval to the next, food predicts nothing and there is no reason for the animal to deviate from a more or less steady rate of responding. (Note that the pause after food on FR schedules is counterproductive: it unnecessarily delays food delivery. More on this later.)

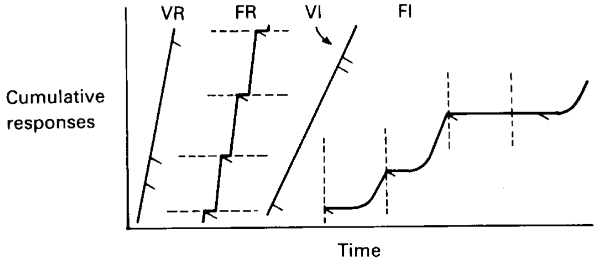


Figure 5.4 Stylized cumulative records of steady-state performance on fixed- and variable-ratio and interval schedules. Rate of responding (responses/time) is represented by the slope of these curves. Dashed lines show when a reinforcer is available for the next response.

The feedback rule for ratio schedules is that reinforcer occurrence depends upon number of responses. There are obviously two other simple possibilities: dependence on *time,* or joint dependence on time and number. Pure dependence on time is an open-loop procedure, in the sense that reinforcer occurrence is then independent of the animal's behavior, so that the response input (labeled *x* in Figure 5.2) doesn't exist. Other names for open-loop procedures are *classical* or *Pavlovian conditioning* (see Chapter 4). I return to them in a moment. The only remaining operant possibility, therefore, is joint control by time and number. The most frequently used procedures of this type are *fixed-* and *variable-interval* schedules. Both require the passage of a certain amount of time followed by a single response for the delivery of the reinforcer. The sequence is important: A response that occurs too early is ineffective; it must occur after the time interval has elapsed.

Figure 5.3 shows how a rat learns to respond on an FR 1 schedule. How might an animal learn to respond on a fixed-interval (Fl) schedule? The process takes much longer than on the simple FR 1, because the animal has to learn about two things: the response contingency — the fact that a response is necessary for each crack at the food — and the minimum interval between food deliveries (i.e., the FI value). He has to learn only the first of these on FR 1. Let's begin with a magazine-trained pigeon maintained at about 80% of its normal body weight (i.e., very hungry!), with the controlling computer set to limit food deliveries to no more than sixty within a single-day session (so as to prevent the animal from gaining weight from day to day). We have trained him to peck (using one of a variety of methods discussed later) but he has so far received food for every effective key peck. Now we introduce him to the fixed-interval procedure, with the interval set to perhaps 60 sec.

Figure 5.5 shows in stylized form the stages that the pigeon's key-pecking goes through as it converges on the final steady-state performance. These stages are by no means clear-cut, nor are the transitions between them perfectly sharp, but we nearly always see these four patterns succeed each other in this order. Each stage takes up many interfood intervals, perhaps 100 or more (i.e., several daily experimental sessions), the number depending on the amount of prior FR 1 training. At first (stage I) each peck-produced food delivery (indicated by the blips in the cumulative record) produces a burst of further pecks, which slowly dies away; the animal pecks slower and slower when these pecks do not result in food. After a time greater than 60 sec has elapsed, and the response rate is now very low, an isolated response immediately produces food and this at once elicits a further pecking burst.

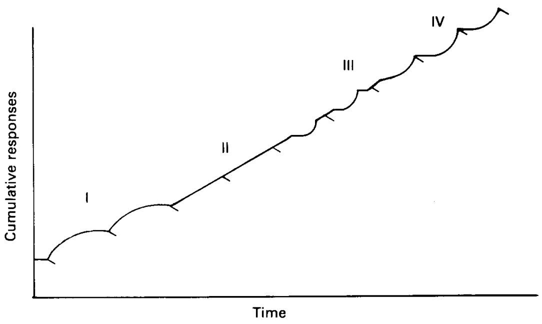


Figure 5.5. Schematic cumulative record of the changing patterns of responding as a pigeon adapts to a fixed-interval schedule. (Adapted from Ferster & Skinner, 1957, p. 117.)

In Stage II, the temporal pattern of pecks between food deliveries changes from negatively accelerated to approximately constant. The pigeon now responds at an approximately steady rate for many intervals. This pattern is succeeded by breaks in the steady responding that take the form of brief periods of acceleration followed by returns to a lower rate of response. This is stage III. This pattern shifts gradually to the final form, which involves a pause in responding after each food delivery, followed by accelerating responding (stage IV). This is the so-called fixed-interval "scallop," a highly reliable pattern shown by many mammals and birds. As Figure 5.4 shows, the steady-state Fl pattern is quite similar to the FR; differences are the lower "running" rate (rate after the postreinforcement pause is over) in FI, the slightly shorter pause (in relation to the typical interfood interval), and the "scalloped" pattern of the FI record, indicating gradual acceleration in responding, rather than the "break-and-run" pattern characteristic of FR.

Each stage of FI acquisition (as this process is termed) makes good adaptive sense. The first stage — a burst of rapid responding after each food delivery — seems to be an innate adaptation to the fact that food often occurs in patches. Finding some food after a lull strongly suggests that there is more where that came from, so that foraging efforts should be stepped up. This is the temporal equivalent of spatial *area-restricted search:* When a pigeon foraging in nature finds some grain after a period of unsuccessful search, his rate of turning increases (that is, he continues to look in the vicinity) and his rate of movement may increase as well. The process is a sort of kinesis, adapted to keep the animal in the "hot" area. An animal on an Fl schedule is restricted to the same "food patch," but it can follow the temporal part of this rule by looking especially hard for more food right after it gets some.

Stage IV is also adaptive: The animal pauses after food because it has learned that no more food is likely for a while and it is free to occupy that time in some other way. The two intervening stages represent the transition period when the animal is gradually giving up its initial, "default" rule (area-restricted search) in favor of a new rule (the Fl scallop) adapted to changed circumstances. Since the animal cannot be certain from day to day that the new pattern of food delivery will persist, it makes sense that he should change only gradually from one behavior to another.

The pause after food on Fl is obviously adaptive, but it is nevertheless usually too short: The efficiency of steady-state Fl performance is surprisingly low. Strictly speaking, only a single response need be made for each food delivery, namely, the first response after 60 sec. Yet a pigeon might make 30-40 key pecks in an average interval, only one of which is essential. Part of the explanation for this inefficiency lies in limitations on the animal's ability to estimate the passage of time, but that is not the whole story. I suggest a possible explanation later.

The difference between Fl and VI parallels that between FR and VR: On VI (if the interreinforcement intervals are chosen randomly) the probability that a response will produce food is constant from moment to moment. Food delivery has no special predictive significance, so that animals tend to respond at a more or less steady rate that is a bit slower than the rate on a comparable VR schedule (I discuss the reason for this in a later chapter).

Equilibrium states

The four patterns shown in Figure 5.4 are the outcome of a converging process. In the acquisition phase, the animal at first shows more or less innate responses to occasional, unpredictable (from its point of view) food deliveries. As food continues to be produced by this interaction, food deliveries begin to take on a predictable pattern; this pattern, in turn, guides future behavior until the process converges to produce the steady-state pattern of behavior shown in the figure. On fixed-interval schedules, for example, the major regularity, the fixed time between successive food deliveries, *depends on the animal' s behavior* as well as on the schedule. If the pigeon pecked slowly or erratically, so that many food deliveries were obtained well after the time at which the apparatus had "set up" (i.e., well after the dashed vertical lines in Figure 5.4), then the time between food deliveries would not be fixed — although the *minimum* interfood interval might still approximate the fixed-interval value. By varying its behavior early in training, the animal is able to detect invariant properties of its hedonic environment: On fixed-interval schedules, the FI value is detected and controls behavior; on FR the ratio value; on VI the mean, minimum interfood interval, and so on.

These examples illustrate a general characteristic of operant behavior: The stimuli that come to control behavior are often themselves dependent on behavior. This kind of interaction is not restricted to the somewhat artificial conditions of the Skinner box. For example, a young squirrel learns about the tastiness of various nuts by first opening them in an exploratory way and sampling the contents. This allows it to learn that some nuts are better than others, so that it will seek out and respond to particularly tasty types that it might otherwise ignore. With additional experience, the animal may come to learn that hazelnuts (say) are to be found in the vicinity of hazel trees or under hazel leaves. Thus at each step, the animal's initial explorations reveal correlations — between the appearance of a nut and its taste, between a habitat and the occurrence of desirable nuts — that guide future behavior. The fixed-interval case is simpler only because the situation is artificially constrained so that the only relevant explorations are along the single dimension of time. The animal varies its distribution of pecks in time and the invariant that emerges is the fixed minimum time between food deliveries. This then guides the future distribution of pecks, which conforms more and more closely to the periodicity of the schedule and, in turn, sharpens the periodicity of food deliveries.

The fixed-interval scallop, and other properties of such stable performances, are aspects of the *equilibrium state* reached by the feedback system illustrated in Figure 5.2. This equilibrium is dependent both on the fixed-interval feedback function, and on the mechanisms that underlie the organism's operant behavior. The examples I have given should make it clear that a particular equilibrium *need not be unique,* however. The final equilibrium depends on two things: the range of *sampling* — the variability in the animal's initial response to the situation, the number of different things it tries; and the *speed of convergence* — the rapidity with which the animal detects emergent regularities and is guided by them. Too little sampling, or too rapid convergence, may mean an equilibrium far from the best possible one. In later chapters I discuss phenomena such as "learned helplessness" and electric-shock-maintained behavior that represent maladaptive equilibria.

Equilibria can be *stable, unstable, neutral,* or *metastable* in response to environmental changes. Stability is not an absolute property of a state but a label for he observed effect of a perturbation. A state may be recoverable following a small perturbation but not after a large one. For example, the physics of soap bubbles shows that their form is the one with the lowest free energy for the number of surfaces: Thus, a spherical bubble returns to its original, efficient shape after a slight deformation. The spherical shape is a stable equilibrium under moderate perturbations. A more drastic deformation will tear the surface, however, whereupon the bubble collapses to a drop and the original state cannot be recovered. The spherical shape is stable under slight deformation, but metastable in response to a severe deformation.

The four types of equilibrium can be illustrated by a visual metaphor. Imagine hat the state of our hypothetical system is represented by the position of a ball in terrain of hills, valleys and plains. On the plain, the ball stays where it is placed: This is neutral equilibrium. In a valley, if the ball isn't moved too far, it returns to the valley floor: This is stable equilibrium. On a hill, even a small displacement lets the ball run down to the valley: This is unstable equilibrium. If the ball starts in a valley it returns to the valley only if it isn't moved too far up the hill; too big a displacement, and it rolls into the next valley: This is metastable equilibrium.

The valley metaphor is realistic in one other way, in that it implies oscillatory behavior as the ball rolls from one valley wall to another after a perturbation. Many natural systems show persistent oscillations in the steady state; motor tracking and predator-prey interactions are well-known examples.

Equilibria on simple reinforcement schedules are generally stable: A given schedule usually yields the same pattern of behavior, and this pattern can usually be recovered after some intervening procedure. Exceptions are of two kinds. Occasionally a pattern is unstable in the sense that it persists for a brief period (which may be several days, or even weeks), but then without any change in the schedule it alters in an irreversible way. For example, occasionally an animal will show instead of the typical scallop pattern a more or less steady rate of responding on a fixed-interval schedule. Indeed, all animals pass through such a period. However, once this pattern changes to the scalloped one, the original pattern never reappears. It, therefore, represents an unstable equilibrium.

A more interesting exception is *metastability.* In this case, the pattern of behavior under a given schedule remains stable, but it is not *recoverable* when the schedule is reimposed after an intervening treatment. This effect is quite common. So-called *spaced responding* provides an example. Hungry pigeons can be trained to space their key pecks in time by delivering food only if a key-peck is separated from the preceding one by at least *t* sec, where *t* is on the order of 10 or 20. They adapt to this procedure with difficulty because the better they adapt, the more frequently they receive food, and the more frequently they get food the more inclined they are to peck. Since more frequent pecking reduces the rate of access to food, the stage is set for a very slow and oscillatory process of adaptation. Pigeons do settle down eventually, however, but at first their performance is far from optimal. For example, an animal initially exposed to a 10-sec timing requirement may after several weeks still space most of its pecks less than 5 sec apart, and the mode of the inter-peck interval distribution may be at only 1 or 2 sec. If the spacing requirement is changed, then, on returning to the original requirement, the pigeon will do much better than before: The average spacing between pecks will be greater and the modal peck closer to the timing requirement. As the animal is exposed to different timing requirements, the performance at each requirement gets better, until the mean and modal inter-peck times come to approximate the spacing requirement. This pattern represents the stable equilibrium. The earlier patterns, in which mean and modal inter-peck time were much shorter than the timing requirement, are metastable equilibria.[6]

The procedures of fixed-ratio, fixed-interval, and variable-interval schedules are summarized in the form of event diagrams in the bottom half of Figure 5.6. The top half shows open-loop (classical conditioning) procedures, which I discuss next.

Classical conditioning

The study of classical conditioning begins with the Russian I. P. Pavlov, whose work made its major impact in the West with the publication in 1927 of an English translation of his lectures on conditioned reflexes. The lectures had been given three years earlier to the Petrograd Military Medical Academy and summarized several decades of active work by a large research group. The subtitle of *Conditioned Reflexes* is "An Investigation of the Physiological Activity of the Cerebral Cortex," which gives a clue to Pavlov's objectives. As a physiologist he was interested in behavior as a tool for understanding the functioning of the brain. Like Sherrington, however, his experiments involved little surgical intervention. Most were purely behavioral, and though, like Sherrington, he often interpreted his results physiologically — inferring waves of excitation and inhibition spreading across the cortex, for example — unlike Sherrington, later - physiological work has not supported his conjectures. Nevertheless, Pavlov's observations, and his theoretical terms, continue to be influential.

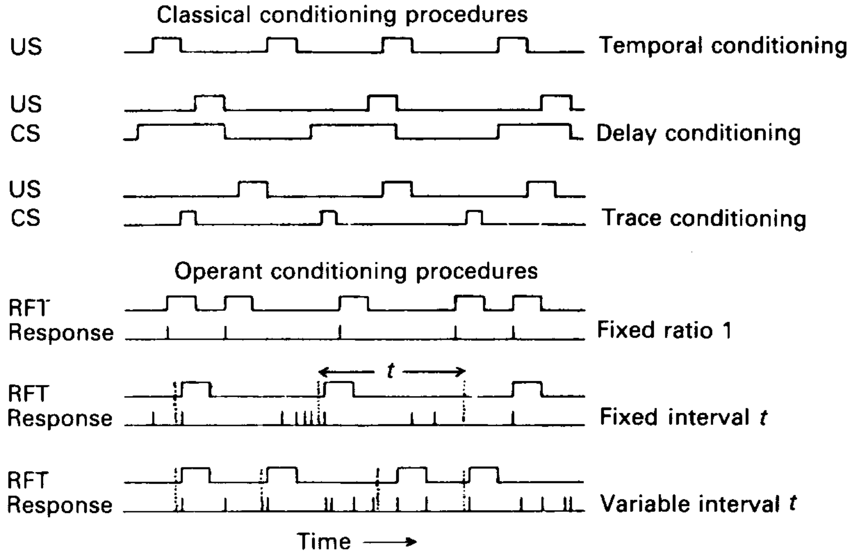


Figure 5.6. Event diagrams of common classical- and operant-conditioning procedures. Time is on the horizontal axis and the occurrence of the labeled event (CS, US, response, RFT = reinforcement) is indicated by upward deflection of the line. In the classical procedures, the occurrence of the US (unconditioned stimulus) is independent of responding, but systematically related to a CS (conditioned stimulus), post-US time or post-CS time. In the operant procedures, the reinforcer depends on a response and (on all schedules other than fixed-ratio 1) on some other time, response or stimulus condition. Thus, in fixed interval, a certain time, *t*, must elapse before a response is reinforced; on fixed-ratio N, N-1 responses must occur before the Nth response is effective.

Pavlov's basic procedure was the one labeled *delay conditioning* in Figure 5.6: A brief stimulus, of a few seconds' duration, such as a tone or a bell or a flashing light, is periodically presented to a dog restrained in a harness. The dog has recovered from a minor operation in which the duct of its salivary gland has been brought to the outside of the cheek, so that the saliva can be collected and measured. At the end of this brief stimulus (which is to become the *conditioned stimulus,* CS for short) some food powder (the *unconditioned stimulus:* US) is placed in the animal's mouth. The food powder, of course, induces salivation; this is the *unconditioned response* (UR). This sequence of operations and effects, tone→ food →salivation, is repeated several times and soon there is a new effect. Salivation now begins to occur when the tone comes on, and before food has actually been placed in the animal's mouth. This is termed the *conditioned* or *conditional response* (CR) — conditional because it depends upon the relation between CS and US during prior training.

This effect can hardly be called startling and must have been observed by anyone who had kept animals and fed them on some kind of schedule. George Bernard Shaw in his satire "The Adventures of the Black Girl in Her Search for God," parodies Pavlov thus:

> "This remarkable discovery cost me twenty-five years of devoted research, during which I cut out the brains of innumerable dogs, and observed their spittle by making holes in their cheeks for them to salivate through.. . The whole scientific world is prostrate at my feet in admiration of this colossal achievement and gratitude for the light it has shed on the great problem of human conduct."
>
> " Why didn't you ask me?" said the black girl. "I could have told you in twenty-five seconds without hurting those poor dogs."
>
> "Your ignorance and presumption are unspeakable," said the old myop. "The fact was known of course to every child: but it had never been proved experimentally in the laboratory; and therefore it was not scientifically known at all. It reached me as an unskilled conjecture: I handed it on as science." (1946, p. 36)

Shaw's parable is a reminder that selling as behavioral "science" the commonplace embedded in jargon is nothing new. But in this case his criticism is not just. Pavlov's contribution was not the discovery of anticipatory salivation, but its measurement and use as a tool to study behavioral processes (the "physiology of the cerebral cortex" in his terminology). For example, he found that if the duration of the CS is longer than a few seconds, the saliva does not begin to flow at once, but is delayed until just before the delivery of food. (Pavlov called this *inhibition of delay.)* This period of delay is fragile, however, in the sense that any unexpected change in the situation immediately causes a copious flow of saliva. This phenomenon, a sort of dishabituation (see Chapter 4), is found in operant as well as classical conditioning; and its study has revealed important things about memory and the organization of action. I have already discussed inhibition and the related phenomenon of successive induction, both of which Pavlov defined and demonstrated. By looking at the effects of varying the type of CS and the time between CS and US, by pairing one stimulus with the US and another with its absence, by looking at what happened when the CS was presented without the US, and by numerous other similar manipulations, Pavlov was able to develop for learned behavior a set of principles comparable to those earlier derived by Sherrington in his studies of spinal reflexes.

Thorndike looked at instrumental learning and decided that its essential feature was temporal contiguity between response and reinforcer ("satisfier"). Pavlov came to the same conclusion about classical conditioning: The necessary and sufficient condition for conditioning is temporal contiguity between CS and US. Both men had good reasons for their belief — although both were wrong. The first experiments that showed why, and point to a sort of alternative, are classical-conditioning experiments.

Pavlov could point to much data suggesting the importance of contiguity. For example, compare delay conditioning with so-called *trace conditioning* (the second and third panels in Figure 5.6). Both procedures involve a delay between CS onset and occurrence of the US, but in the trace case, the CS ends some time before the US begins. Trace conditioning is much more difficult to get than delayed conditioning, which suggests that any delay between CS and US is detrimental to conditioning. Many other experiments have demonstrated the bad effects on conditioning of CS-US delays. The only discordant note was provided by *temporal conditioning,* the top panel in Figure 5.6, which is just periodic presentation of the US (like a fixed-interval schedule, but without the response requirement: Temporal conditioning is also known as a *fixed-time schedule).* In temporal conditioning, the US is also a (temporal) CS, like the neutral CS in trace conditioning. Despite the delay between CS (US) and US, temporal conditioning is very effective, even with long delays. This difference between temporal and trace conditioning seems to depend on the properties of memory — which also account for other examples of long-delay conditioning discovered subsequently (see Chapters 12 and 13). But temporal conditioning attracted little attention until relatively recently, and the main attack on contiguity came from another quarter.

CONTINGENCY AND FEEDBACK FUNCTIONS

A seminal paper in 1967 by R. A. Rescorla made a major advance. Rescorla used a classical conditioning procedure, invented by W. K. Estes and B. F. Skinner in 1943, that does not involve salivation at all. The procedure is in fact a mixture of both operant- and classical-conditioning procedures. The key ingredient is a variable-interval schedule of food reinforcement. Responding on VI schedules is an admirable *baseline* with which to study the effects of other independent variables: Other things being equal, the animal responds at a steady rate; hence, any change in rate associated with the presentation of a stimulus can safely be attributed to the stimulus, rather than to accidental variation.

Estes and Skinner made use of a VI baseline to study the effect of occasionally presenting a relatively brief, neutral stimulus of the type used by Pavlov. After an initial "novelty" effect, the animal (usually a rat) continues to respond when the stimulus is present at about the same rate as when it is absent. This is the control condition, which establishes that the stimulus by itself has no effect. In the next phase, the rat is briefly shocked (through the metal grid floor) at some time during the stimulus presentation. After a few such stimulus-shock pairings, the stimulus produces a clearly recognizable suppression of lever pressing, as shown in Figure 5.7. This *conditioned suppression* (also termed the *conditioned emotional response,* or CER) can be measured by the relative rate of responding in the presence of the CS, the *suppression ratio:* $S = N_s/(N_s + N_{ns})$, where S is the suppression ratio, N_s is the number of lever presses when the stimulus is present, and N_{ns} is the number of lever presses during a comparable period when the stimulus is absent. $S = 0$ if the animal stops pressing completely in the presence of the CS, and .5 if the CS has no effect.

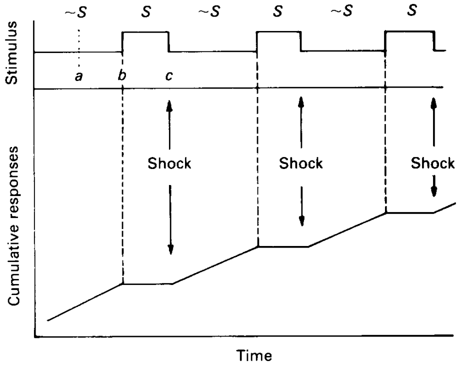
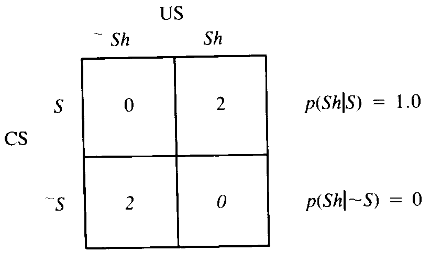


Figure 5.7. The suppression of a response maintained by a variable-interval schedule of reinforcement during a stimulus, *S*, ending with a brief electric shock. The degree of suppression is measured by comparing response rate during the stimulus (period *bc*) with responding during the same period before stimulus onset (period *ab*).

Conditioned suppression behaves in essentially the same way as the salivation of Pavlov's dogs, but has numerous practical advantages. Rats are cheaper than dogs; no operation is required; there are fewer physiological limitations on lever pressing than on salivation; it is less messy. Although other classical conditioning methods are sometimes used in Western laboratories (e.g., measurement of skin resistance, of blinking, or of the nictitating-membrane response in rabbits), salivation is hardly studied at all, and the conditioned-suppression method is widely favored.[7]



		US		
		~*Sh*	*Sh*	
CS	*S*	0	2	$p(Sh \mid S) = 1.0$
	~*S*	2	0	$p(Sh \mid \sim S) = 0$

Table 5.1. *Correlated condition.*

A simplified version of Rescorla's procedure is shown in Figure 5.8. There are two stimuli (e.g., a tone vs. a light, or buzzer vs. absence of buzzer), labeled *S* and ~*S* ("not-*S*," the absence of *S*). The stimuli are perhaps 60 s in duration and, in this simplified version, occur in strict alternation, with about 50 such cycles making up a daily experimental session (only 2 cycles are shown in the figure). In the *correlated* condition (second row in Figure 5.8), brief, randomly spaced shocks occur only during stimulus *S*. In the *uncorrelated* condition, the shocks occur throughout the experimental session, that is, indiscriminately in the presence of both *S* and ~*S*. (The uncorrelated condition is sometimes also called the *truly random control* condition.)

The effect of these two procedures on lever-pressing maintained by the variable-interval schedule is shown in the bottom two rows of the figure. In the *correlated* condition animals typically respond for food only in the presence of stimulus ~*S,* which is not associated with shock: This is conditioned-suppression, just discussed. The interesting result is obtained in the *uncorrelated* condition (bottom row): In this condition animals respond indiscriminately in the presence of both stimuli, although at a somewhat lower rate than in stimulus ~*S* in the correlated condition.

	~ *Sh*	*Sh*	
S	0	2	*p*(*Sh*|*S*) = 1.0
~*S*	0	2	*p*(*Sh*|~*S*) = 1.0

Table 5.2. *Random condition.*

	~ *Sh*	*Sh*	
S	2	3	*p*(*Sh*|*S*) = .6
~*S*	*4*	*1*	*p*(*Sh*|~*S*) = .2

Table 5.3. *Partially correlated condition.*

	~ *Sh*	*Sh*	"Instants" in:
S	10	10	*S* = 20
~S	*10*	*0*	~*S* = 10

Table 5.4. *Correlated condition–time bins.*

	~ *Sh*	*Sh*	
S	.5	.5	*p*(*Sh*|*S*) = .5
~*S*	*1.0*	*0*	*p*(*Sh*|~*S*) = 0

Table 5.5. *Correlated condition–conditional probabilities.*

This result completely rules out CS-US contiguity as a sufficient explanation for classical conditioning. Simple pairing of US (shock) and CS (stimulus *S)* cannot be sufficient for conditioning, since this pairing holds in both the *correlated* and *uncorrelated* conditions of Rescorla's experiment; yet conditioning occurred only in the *correlated* condition. What, then, are the necessary and sufficient conditions for classical conditioning?

Intuitively, the answer is clear. The animals show conditioning to a stimulus only when it *predicts* the US: CS and US must therefore be correlated for conditioning to occur. This conclusion is appealing and widely accepted, but deceptive: One might say of the concept of "predictability," as of the bikini: What it reveals is suggestive, but what it conceals is vital. "Predictability" is something quite different from "contiguity." Contiguity is a perfectly unambiguous, quantitative time relation between two events. But the predictability of something depends upon the *knowledge* of the observer: Since Newton, we can predict when Halley's comet will return, whereas before none could do so; we know something Newton's predecessors did not. The situations used to study conditioning in rats are so simple, and our understanding of them so intuitive, that it is hard to define just what is involved in detecting the kinds of correlation depicted in Figure 5.8 — hard even to realize that correlation is not a simple property like weight or duration. Nevertheless, an explanation of conditioning in terms of correlation or predictability is a functional explanation, and therefore in many ways less powerful than Pavlov's completely

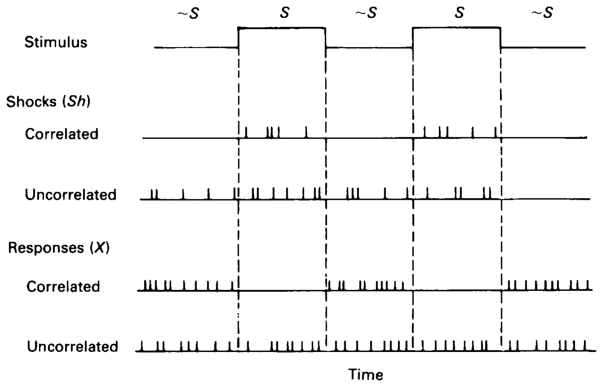


Figure 5.8. Stimulus contingencies in classical conditioning. The top panel shows the alternation of two stimuli. The next two panels show correlated and uncorrelated stimulus contingencies. The last two panels show the effects of each on responding maintained by a VI food schedule.

mechanistic contiguity account. Rescorla's explanation is better than Pavlov's, but its success carries a cost: A gain in comprehensiveness has also meant a loss in precision.

The idea of correlation can be made more precise with the aid of a common device, the *contingency table.* Tables 5.1 and 5.2 are computed for the *correlated* and *uncorrelated* conditions in Figure 5.8. The rows correspond to stimuli *(S* and *~S)* and the columns to the occurrence or nonoccurrence of shock *(Sh* and *~Sh).* Thus, the entry in the upper right cell in Table 5.1 is the number of occurrences of stimulus *S* when at least one shock occurred (both presentations of *S* are accompanied by shock in Figure 5.8). The bottom right cell gives the number of times when *~S* occurred and was accompanied by shock (zero), and so on. The concept of a *stimulus contingency* is obvious from the comparison of Tables 5.1 and 5.2: When the presence or absence of the stimulus is a predictor of shock, entries in the major diagonal of the table are high and entries elsewhere are low, as in Table 5.1. When the presence or absence of the stimulus is uncorrelated with the presence or absence of shock, the rows are the same, as in Table 5.2. The entries to the right of the tables *(P(Sh /S),* etc.) are conditional probabilities, which are defined in a moment.

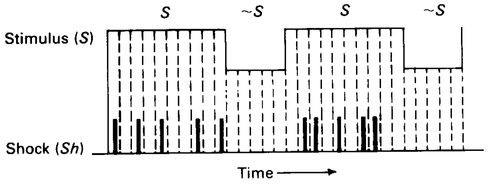


Figure 5.9. Analysis of stimulus contingencies by time bins.

Tables 5.1 and 5.2 present "pure cases": The contingency between S and shock is perfect in Table 5.1 and zero in Table 5.2. Intermediate cases are also possible, and one is illustrated in Table 5.3. There is some correlation between *S* and shock in Table 5.3, but it is clearly less than perfect. The degree of association between shock and stimulus in these tables can be quantified using the X^2 statistic, or the contingency coefficient, ϕ which is just (X^2/N), where *N* is the total number of cell entries.[8] I describe a graphical method for representing the degree of contingency in a moment.

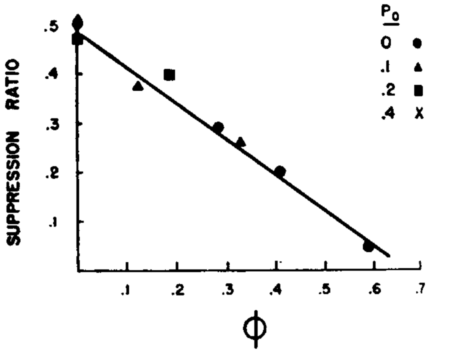


Figure 5.10. Suppression ratio as a function of ϕ for experimental results from Rescorla (1968). Suppression ratio is response rate in the CS divided by rate in the CS plus rate in its absence (" no suppression" = ratio of .5; which is the value for both noncontingent points, where ϕ = 0). Increasing suppression is indicated by smaller suppression values. (From Gibbon, Berryman, & Thompson, 1974, Figure 3.)

The contingency measures represented by Tables 5.1-5.3 are simplified in an important way: They take no account of the relative durations of *S* and *~S.* For example, even if shocks occur independently of *S,* if *S* is on for, say, 2/3 of the time and *~S* for only 1/3, then the column entries for *Sh* in the table would be higher for *S* than *~S.* This gives the appearance of a contingency between *S* and *Sh* even though none exists.

One way of handling this difficulty is shown in Figure 5.9. Time is divided up into discrete intervals small enough so that no more than one shock can occur in an interval. The total number of entries in the contingency table, *N,* is then the total number of such " instants." Cell entries are the number of instances in *S* and *~S* when *Sh* or *~Sh* occurred. Table 5.4 shows some hypothetical numbers for the *correlated* condition in an experiment in which *S* was twice as long as *~S,* and shock occurred in just half the instants in *S.* Table 5.5 is Table 5.4 reduced to a standard form that takes account of the different durations of *S* and *~S.* Each cell entry is the *conditional probability* of shock, given that *S* or *~S* is occurring (i.e., *p(Sh/S)* and *p(Sh/~S));* that is, each cell entry in Table 5.5 is just the entries in Table 5.4 divided by the row totals.[9]

The amount of conditioning to a stimulus in a classical-conditioning situation is directly related to the degree of contingency between the conditioned stimulus and the unconditioned stimulus. Figure 5.10 shows a particularly neat example of this relation: Data from an experiment by Rescorla show that the suppression ratio is linearly related, with negative slope, to the value of ϕ, the coefficient of contingency between CS and US. Since low suppression ratios mean high suppression, these data show the direct relation between the effect of a stimulus and its correlation with shock.[10]

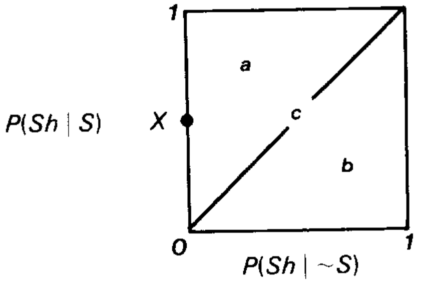


Figure 5.11. Stimulus-stimulus (CS-US) contingency space.

Contingency space

Table 5.5 is the basis for the general, two-dimensional, *contingency space* shown as Figure 5.11. Since the row totals always add to 1.0, a given contingency table can be represented by just one of the two columns. By convention the righthand column is usually chosen, that is, *p(Sh|S)* and *p(Sh|~S).* Thus, each such contingency table defines one point in contingency space. (Table 5.5 is represented by the point labeled "X" on the ordinate.)

The contingency space is divided into three regions: (a) Above and to the left of he major diagonal is the region of *positive contingencies,* where *p(Sh|S) > p(Sh|~S)* (shocks are more likely in *S).* (b) Below and to the right of the major diagonal is the region of *negative contingencies,* where *p(Sh|S) < p(Sh|~S)* (shocks are less likely in *S).* (c) The major diagonal itself defines he *absence* of contingency between *Sh* and *S,* where *p(Sh|S) = p(Sh|~S)* (this is the uncorrelated condition: Shocks are equally likely in *S* and *S).*

Positive contingencies generally produce *excitatory conditioning,* that is, the contingent stimulus produces an effect of the same sort as the US - suppression of food-reinforced responding in the CER experiment. Negative contingencies generally produce *inhibitory conditioning,* that is, effects of a sort opposite to those of the unconditioned stimulus. For example, imagine a CER experiment in which shocks occasionally occur on the baseline (i.e., in the absence of any CS). We could present two kinds of CS: a stimulus in which no shocks occur (a "safety signal" — inhibitory CS), and a signal in which the shock rate is higher than baseline (a "warning signal" — excitatory CS). The safety signal will produce an *increase* in lever pressing (suppression ratio > .5, an inhibitory effect in this context), whereas the shock-correlated CS will produce suppression relative to baseline (suppression ratio < .5, an excitatory effect in this context).

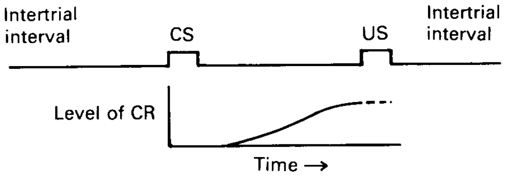


Figure 5.12. Typical time relations between conditioned response (CR) and trace-conditioned stimulus (CS) in salivary conditioning.

Temporal and trace conditioning

For simplicity, and historical reasons, I have introduced the notion of contingency in connection with the CER procedure and synchronous stimuli — stimuli that occur at the same time as the shocks they signal. But a similar analysis can be applied to temporal stimuli, as in temporal and trace conditioning. These temporal procedures raise two questions: (a) What is the effect of the trace CS on the conditioned response (CR)? Is the CR more or less likely to occur when the CS occurs compared to when the CS is omitted? (This is the question just discussed in connection with the CER procedure.) (b) When does the CR occur? Is it during the CS, immediately after, or after a delay? (This is a question specific to temporal procedures.)

The answer to the how-likely-is-the-CR question is " it depends" — on the temporal relations among CS, US, and other stimuli in the situation. The timing of the CR also depends on these factors to some extent, although there are some general things one can say: The CR almost never occurs during a trace CS, always afterward. It is also often delayed, more or less in proportion to the CS-US interval: the longer the CS-US interval, the longer the trace CR is delayed after CS offset. Figure 5.12 shows in stylized form the typical time course of a trace-conditioned response.

I've already mentioned that trace conditioning is hard to get. To begin to see why, look at the four conditioning sequences diagrammed in Figure 5.13. The sequences are typical classical-conditioning procedures, consisting of an *intertrial interval,* the period between the end of the US and the onset of the next CS (period US-CS) and the *trial period* between the CS and the ensuing US (period CS-US). In the diagram these two periods add to a constant, the US-US interval. I'll consider this case first, then look at what should happen if the intertrial interval is allowed to vary while the CS-US period remains constant. These two periods are formally analogous to the ~*S* and *S* periods in the CER procedure we just looked at. They may not be analogous from the animal's point of view, of course, because they depend upon his ability to *remember* the event initiating the period — the CER procedure imposes no such memory requirement, because the stimuli are continuously present. The behavior to be expected from these four procedures depends entirely on how the animal uses the temporal information available to him.

What is the effect of the CS on the probability of a CR in these procedures? When does trace conditioning occur and when does it fail? Consider two possibilities: (a) The animal has perfect memory (i.e., can use either the CS, the US, or both as time markers). (b) The animal's memory is imperfect.

In the first case, since the CS is always closer than anything else to the next US, the CS might be said to predict the US, and trace conditioning should occur. Since the times between US and US and between CS and US are both fixed, however, the animal might use either or both as time markers. We know that temporal conditioning, sequence D in the figure, yields excellent conditioning, so we know that animals can use the information provided by a fixed US-US interval. If he uses the US, then the CS will have no special effect, and trace conditioning will not occur. But since the accuracy with which an animal can tell time is roughly proportional to the time interval involved (Weber's law for time; see Chapter 12), the animal will obviously do better to use the CS, the stimulus closest to the next US. Moreover, this advantage will be greater the closer the CS to the US: Sequence A should therefore produce the most reliable trace conditioning, B next, and C the worst. If the animal can use either CS or US as a trace stimulus, he should always use the CS.

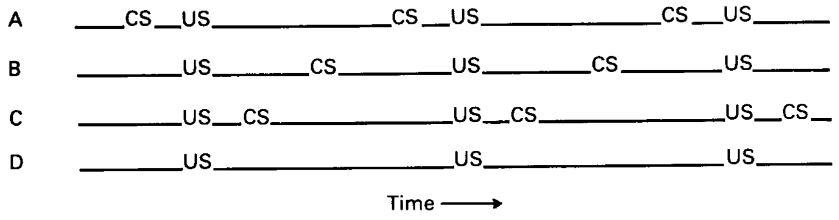


Figure 5.13. Three CS placements in trace conditioning (A, B, and C), compared with temporal conditioning (D).

If the animal's memory is imperfect, however, then some events may make better time markers than others. In particular, for reasons I discuss in Chapter 12, the US may be a much better time marker than a "neutral" CS. In this case we have two factors that act in opposite directions: The CS is always closer than anything else to the US. Hence, trace conditioning to the CS is favored over trace (temporal) conditioning to the US. But if the US is better remembered (makes a better time marker) than the CS, then, other things being equal, temporal conditioning will be favored over trace conditioning. Trace conditioning should, therefore, occur only under two conditions: (a) When the CS-US interval is much shorter than the US-US interval; or (b) when the US-US interval is *variable,* so that post-US time cannot be used to predict US occurrence. Both these predictions have generally been supported.

Obviously animals behave adaptively in classical-conditioning experiments, much more so than the earlier notion of automatic conditioning-by-contiguity suggests. For the most part, animals become conditioned to the stimulus that predicts the US, and apparent failures to do so in trace conditioning seem to reflect special attention to the US, which is undoubtedly adaptive in other contexts. The subtlety of this behavior poses considerable problems for theory, however. So long as simple pairing seemed to be the critical operation for classical conditioning, attention could be focused on procedural details — the relation between conditioned and unconditioned responses, the effects of CS-US delays, the effects of CS salience, and so on — with the assurance that the basic mechanism was known. It was not. We still know little about the computational process that allows animals to identify and react just to those aspects of their environment that predict hedonic events. I return to this problem in Chapter 13.

Response contingencies and feedback functions

Both Pavlov and Thorndike thought that contiguity was the only, or at least the major, factor in learning. Later experiments have shown that contiguity is not sufficient, that animals are able to detect correlations between CS and US in classical-conditioning situations. As one might expect, the same thing has turned out to be true of operant conditioning — although the theoretical analysis is a bit more complicated.

The problem for an animal in a classical-conditioning experiment is to detect which stimuli predict the US, and how well they predict it. The problem in an operant-conditioning experiment is similar, namely, what aspects of the animal's *behavior* (i.e., what responses) predict the US (reinforcer), and how well do they predict it. Both are examples of what is called the *assignment-of-credit* problem. The only difference between the problems of stimulus and response selection is that the animal can control its own behavior, whereas it cannot control the CS in a classical-conditioning experiment. Hence, an animal in an operant-conditioning experiment need not take at face value a given correlation between its behavior and reinforcer occurrence: It can increase or decrease its response rate, or vary responding in other ways, and see if the correlation still holds up. This difference means that the animal can hope to arrive at the *cause(s)* of the reinforcer in an operant-conditioning experiment, but is limited to detecting *correlations* between CS and US in classical conditioning experiments.

We can approach the response-selection problem in two ways: One is by an extension of the discrete-time analysis just applied to CER conditioning, the other by an extension of the memory analysis applied to trace conditioning. I discuss the discrete-time method here, leaving more complex analysis to a later chapter.

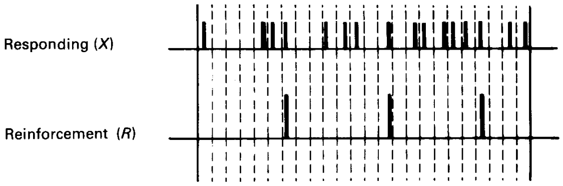


Figure 5.14. Analysis of response-reinforcement contingency by time bins.

The discrete-time representation of contingency in Figure 5.9 can be applied directly to response selection in operant conditioning in the following way. Figure 5.14 shows two event records: responding at the top, and associated variable-interval reinforcement on the bottom. As before, time is divided into discrete "instants" small enough that no more than one response, or reinforcer, can occur in each. To find out if the response predicts the reinforcer we can ask: How many times is a reinforcer accompanied by a response (i.e., both in the same instant)? How many times does it occur without a response? How many times does a response occur unaccompanied by a reinforcer? How many times does neither response nor reinforcer occur?

The contingency Table 5.6 shows the answers for the sample of responses and reinforcements shown in Figure 5.14. Table 5.7 shows the same data in conditional-probability form: The

		Reinforcer	
		~ *R*	*R*
Response	*x*	13	3
	~*x*	8	0

N = 24

Table 5.6. *Response–contingent reinforcement.*

	~ *R*	*R*
x	.81	.19
~*x*	1.0	0

Table 5.7. *Response-contingent reinforcement–conditional probabilities.*

	~ *R*	*R*
x	x – *R*(*x*)	*R*(*x*)
~*x*	1 – x	0

Table 5.8. *Response-contingent reinforcement–analysis by rates.*

entries in Table 5.7 are the entries in Table 5.6 divided by the column totals. They show the conditional probabilities that *R* or ~*R* will occur, given *x* or ~*x*. An obvious difference between the contingency tables for response and stimulus selection is that the response table is partly under the animal's control. For example, if the animal responds more slowly, the entries in the upper-left cell (unsuccessful responses) will decrease much more rapidly than entries in the upper right cell (reinforced responses); that is, responding predicts reinforcement much better — the contingency between the two improves — as response rate decreases. This follows from the properties of a variable-interval schedule: The more slowly the animal responds, the more likely that each response will produce a reinforcer. With further decreases in response rate, the entries in the upper right cell will also begin to decrease. As response rate decreases to zero, entries in all save the lower left cell vanish.

The lower right cell in Table 5.6 is always zero for the standard reinforcement schedules — the animal never gets a reinforcer unless he responds. Consequently variations in the animal's rate of responding simply move the point representing the matrix up and down the vertical axis of the $p(R \mid \sim x)$ versus $p(R \mid x)$ contingency space (See Figure 5.16). Molar *feedback functions,* which show the relation between the rates of responding and reinforcement, are therefore more useful than contingency tables as representations of standard schedules.

The relationship between the feedback function and a contingency table such as Table 5.6 can be easily seen by considering the number of responses, *x,* and reinforcements for those responses, *R*(*x*), over a unit time period (i.e., a period chosen so that *x* and *R*(*x*) are just rates). Table 5.6 can then be rewritten entirely in terms of *x* and *R*(*x*), as shown in Table 5.8. The feedback function is simply the systematic relation between *x* and *R*(*x*) enforced by the schedule. Each point on the feedback function (i.e., each pair of [*x, R*(*x*)] values), therefore, defines a separate contingency table.

Feedback functions for common schedules

Molar feedback functions are important for topics discussed later in the book. To make the concept as clear as possible, and as a way of introducing some new procedures, I now show how to derive feedback functions for some common schedules.

Ratio schedules. Ratio schedules prescribe either a fixed (FR) or variable (but with fixed mean: VR) number of responses per reinforcement. The molar feedback function makes no distinction between FR and VR, since it simply relates aggregate responding to aggregate reinforcement. This function, derived earlier, is a simple proportion: *R*(*x*) = *x/M,* where *R*(*x*) is reinforcement rate, *x* is response rate, and *M* is the ratio value. The ratio feedback function is therefore a straight line through the origin (see Figure 5.15).

Interval schedules. Fixed-interval schedules must be treated differently from variable-interval, because the animal can predict when food will be available on FI almost perfectly, whereas VI schedules are explicitly designed to prevent this. I just consider VI here. For a VI schedule with random interreinforcement intervals, the average time between reinforcements is made up of two times: the prescribed minimum interreinforcement interval (the VI value), which can be written as $1/a$, where a is the maximum possible reinforcement *rate,* and d, which is the delay between the time when reinforcement is available for a response and the time when the next response actually occurs. Thus,

$$D(x) = 1/R(x) = 1/a + d, \qquad (5.2)$$

where $D(x)$ is the actual mean time between reinforcements and $R(x)$ the obtained rate of reinforcement, as before. If we specify the temporal pattern of responding, then d can be expressed as a function of x, the average rate of responding. In the simplest case, if responding is random in time, then

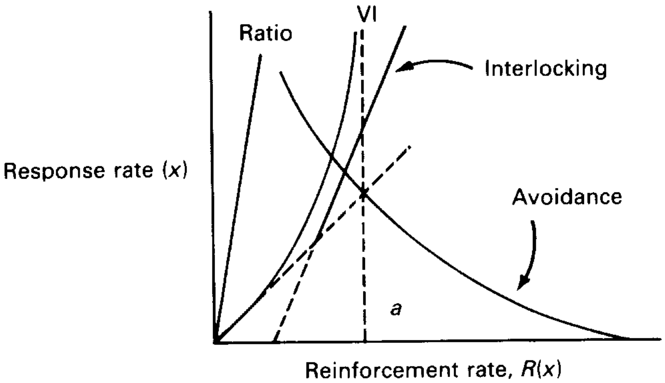


Figure 5.15. Molar feedback functions for four operant schedules: ratio, variable-interval, interlocking, and avoidance (shock postponement). These curves show how the schedule enforces as relation between response rate (independent variable: x) and reinforcement (or punishment, for avoidance) rate: (dependent variable: $R(x)$). The usual convention would put $R(x)$ on the vertical axis and x on the horizontal, The axes are reversed here because in all subsequent discussion we will be much more interested in the reverse relation, where $R(x)$ is the independent variable and x the dependent variable. For consistency, I adopt throughout the latter arrangement.

$$d = 1/x, \qquad (5.3)$$

the expected time from reinforcement setup to a response is just the reciprocal of the average response rate. Combining Equations 5.2 and 5.3 yields the actual feedback function, which is therefore

$$R(x) = ax/(a + x), \qquad (5.4)$$

a negatively accelerated function which tends to a as $x \rightarrow \infty$, and to x as $x \rightarrow 0$; that is, it has asymptotes at $R(x) = a$, and $R(x) = x$, as shown in Figure 5.15.

If responding is not random, or if it depends on postreinforcement time (as in fixed-interval schedules), the delay, d, may depend on properties of responding in addition to rate. In this case the properties of the schedule cannot be captured by a simple response rate versus reinforcement rate feedback function; something more complicated is required. This difficulty simply emphasizes that although the concept of a feedback function is perfectly general, to get one in a form simple enough to be useful may mean making quite drastic simplifying assumptions about behavior.

Interlocking schedules. Interlocking schedules combine ratio and interval features. Reinforcement is delivered for a response once a weighted sum of time and number of responses exceeds a fixed value. If we neglect the final response requirement, this schedule can be represented by the relation

$$aN(x) + bt = 1, \qquad (5.5)$$

where $N(x)$ is the number of responses made since the last reinforcement, t is the time since the last reinforcement, and a and b are positive constants. But $t = 1/R(x)$ and $N(x)/t = x$, the rate of responding; Hence Equation 5.5 reduces to the feedback function

$$R(x) = ax + b, \qquad (5.6)$$

which is linear, like the ratio function. The line doesn't go all the way to the $R(x)$ axis because at least one response is required for each reinforcer: Hence the interlocking schedule feedback function must have the FR 1 line as an asymptote. Figure 5.15 shows the interlocking schedule function as two line segments: Equation 5.6, and the FR 1 function $R(x) = x$.

These three schedules, ratio, VI, and interlocking, are all examples of positive contingencies: Reinforcement rate increases with response rate. Positive contingencies are obviously appropriate for use with positive reinforcers. Conversely, escape, avoidance, and omission schedules all involve negative contingencies: "Reinforcement" (actually, punishment, delivery of an aversive stimulus) rate decreases as response rate increases. Negative contingencies are appropriate for negative reinforcers[1] (as earlier defined in terms of approach and withdrawal). In general, the product of the signs of contingency and reinforcer must be positive if behavior is to be sustained. This rule is true of "strong" reinforcers like food for a highly deprived animal, or electric shock, which usually affect behavior in a consistent way. Animals will cease to respond if responding produces shock (this is the procedure of *punishment*) or if responding prevents food. There are some exceptions, however. As I explain in Chapter 7, it is possible to have too much of a good thing, and some things are reinforcing only in certain quantities (in the limit, this is even true of food, though not, perhaps, of money and other human goodies). Even if the reinforcer maintains its value, paradoxical effects can be produced under special conditions: Animals will respond to produce electric shock; and hungry pigeons can be induced to peck even if pecking prevents food delivery. These exceptions are of interest for the light they shed on the mechanisms of operant behavior, and I deal with them in more detail later.

Escape, avoidance, and omission schedules. None of these negative contingencies lends itself easily to algebraic treatment.[11] In escape, omission, and so called discriminated avoidance procedures, the opportunity to respond is usually restricted to discrete trials, separated by an intertrial interval. For example, omission training and discriminated avoidance resemble classical delay conditioning, with the added feature that a response during the CS eliminates the US on that trial. (The only difference between omission training and discriminated avoidance is that the former term is used when the US is a positive reinforcer, the latter when it is a negative reinforcer.) Escape is also a discrete-trial procedure: An aversive stimulus such as shock or a loud noise is continuously present during a trial and a response eliminates it. The contingencies in all these cases are obviously negative ones, but the discrete nature of the procedures, and the addition of a trial stimulus, complicates a feedback-function analysis.

Shock postponement (also termed *unsignaled* or *Sidman avoidance)* is a continuous procedure that is very effective in maintaining operant behavior and does allow a relatively simple analysis. In its standard form, the procedure presents brief shocks to the animal at fixed intervals of time, the shock-shock (*S*S*) interval. If the animal responds at any time, the next shock is delayed by a fixed interval, the response-shock (*R*S*) interval. For example, the *S*S* interval might be 20 s and the *R*S* interval 10 s. Obviously if the animal is careful to respond at least once every 10 s, it need never receive shock. As response rate declines more shocks are received, their number depending upon the distribution of interresponse times (IRTs). If the distribution is sharply peaked, with most IRTs close to the mean value, then even small shifts in the mean will produce a substantial increase in the frequency of shock. Conversely, if the distribution of IRTs is quite broad, a decrease in response rate may result in quite a gradual increase in shock rate. The feedback function therefore depends on the IRT *distribution* as well as the rate of responding.

Most animals that respond at all on shock-postponement schedules respond at a substantial rate. Consequently, of the few shocks they receive, most are determined by the *R*S* interval, very few by the *S*S* interval. (Animals usually respond immediately after a shock, which also

[1] The original label for aversive stimuli was "punishers"; the term negative reinforcement was reserved for schedules in which an aversive event was delayed or withdrawn. In this scheme, both positive and negative reinforcers have positive effects on behavior. But, alas, the term negative reinforcement is now frequently used for the aversive events themselves, so that positive reinforcers enhance behavior and negative reinforcers suppress it. See endnote 3.

tends to eliminate the *S*S* interval as a factor in the maintenance of this behavior — although not, of course, in its acquisition.) If the IRT distribution is known, it is therefore possible to compute the number of IRTs longer than the *R*S* interval as a function of the mean IRT. Converting these values to rates of responding and shock then yields the feedback function.

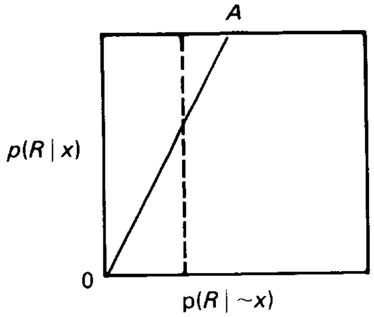


Figure 5.16. Response-reinforcement contingency space.

On first exposure to shock postponement (as to any schedule), responding is likely to be approximately random in time. Later the interresponse-time distribution is likely to become more sharply peaked, approximately at the postponement (*R*S*) value, *T.* For the random case, and *R*S* = *S*S* = *T,* the feedback function is

$$R(x) = x.\exp(-xT)/(1 - \exp(-xT)),$$

which is shown in Figure 5.15.[12] The function has the expected properties: As response rate increases, reinforcement rate decreases, and when response rate is zero, reinforcement (shock) rate equals 1/T.

These four examples were chosen for their simplicity, to illustrate the concept of a feedback function, to give some idea of the variety of possible functions, and to make clear the difference between positive (increasing) and negative (decreasing) functions.

The detection of response contingency

So far I have considered only conventional reinforcement schedules, in which the reinforcement, when it occurs, always requires a response. What if we relax the assumption that reinforcement must always follow a response? For example, suppose we take a pigeon already trained to respond for food on a VI schedule, and arrange that half the reinforcers are delivered as soon as the VI timer "sets up," independently of responding (this is known as a *mixed variable-interval, variable-time* [mix VI VT] schedule): What effect will this have? Since pigeons peck fast for food on VI schedules, this change in procedure will have little effect on the overall food rate; all we have done is degrade the contingency between food and pecking: The lower-right cell in Tables 5.6-5.8 will not now be zero. Consequently, the pigeon should now peck more slowly than before, and indeed this is the usual result in such experiments.[13]

How does the animal detect the degradation in contingency? The answer to this is not fully known, but we can get some idea of what is involved by looking at changes in molar contingencies, and changes in temporal (contiguity) relations.

The change in molar response contingency can be represented in a contingency space, as shown in Figure 5.16. The location of the point representing degree of contingency depends on response rate: When rate is high, the point will be close to the diagonal, because many response-reinforcer conjunctions will be accidental. Since half the reinforcers are response dependent, *p(R|x)* can never be less than twice *p(R|~x),* however. At the other extreme, when response rate is very low, almost every response is reinforced, so that *p(R | x)* → 1. *p(R | ~x)* is fixed by the variable-interval value, hence response rate variation can only move the point up and down the vertical line corresponding to *p(R| x)* — where that line lies on the horizontal axis depends upon the duration of the "instants" (see Figure 5.14) we have chosen. Thus, response-rate variation shifts the contingency up and down the vertical line segment between the upper bound and the line *p*(*Rx*) = 2*p*(*R* | ~*x*) (line OA), as shown in the figure.

Variation in absolute frequency of reinforcement (variable-interval value) and the proportion of reinforcers that are response independent both have identifiable effects in the contingency space. Variation in interval value shifts the location of the *p(R |~ x)* line along the horizontal axis; variation in the proportion of response-independent reinforcers varies the slope of the radial constraint line OA. Thus, the pigeon has ample molar information to detect a change in response contingency.

He also has molecular information, in the form of time differences. The situation is illustrated in Figure 5.17, which shows the differences to be expected when reinforcement is response-independent versus response dependent. Response-reinforcer delay will vary from one reinforcer to the next, but the variation will be less, and the mean value smaller, when the reinforcement is response dependent. When the distributions of d_D and d_I are reduced to a standard form, in which the variances (spreads) are equal, the separation of the means is a measure of the *detectability* of the difference between them. The significant thing about these two distributions is that the animal can separate them as much as he wishes by slowing his response rate: Evidently the best test for response contingency is to slow response rate and see what happens to the distribution of response-reinforcer delays.

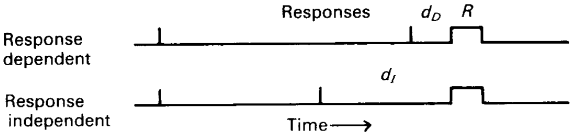


Figure 5.17. Time-delay differences in response-dependent (top) and response-independent (bottom) reinforcement.

Animals are very sensitive to these time-distribution differences. I describe in a later chapter an experiment in which a pigeon had to make a choice depending on whether a just-preceding event had occurred independently of its peck or not — the birds were able to detect very small time differences. Animals seem also to be sensitive to the contingency-detecting properties of a decrease in response rate, and this may account for their problems with spaced-responding schedules — which mimic VI schedules in showing an increasing probability of payoff as response rate decreases. When the spacing requirement is long (> 30 sec or so), pigeons treat spaced-responding schedules just like long VI schedules: They respond at a relatively high rate, and the peak of the IRT distribution is at S sec or less.

Sampling always carries some cost. Responding more slowly helps the pigeon estimate the degree of response contingency, but it also delays food because many response-contingent reinforcers will then be "set up" for some time before they are collected. How should the animal weigh this cost? Figure 5.18 illustrates the problem posed by a fixed-interval schedule. The distribution shows the spread in the animal's estimate of time-to-food, as measured by his postfood pause. Area *A,* to the left of *T* (the FI value), represents unnecessary anticipatory responses, responses made before food is made available by the Fl programmer. Area *B,* to the right of *T,* represents unnecessary delays, interfood intervals longer than necessary because the animal waited to respond until after reinforcement had set up. If the animal shifts *t* (the mode of his pause distribution) to the left, he will reduce delays *(B),* but at the cost of increasing the number of unnecessary responses *(A),* and conversely. As we have already seen, most animals respond too early on FI, which suggests that they weigh unnecessary delays much more than unnecessary responses. Or, to put the same thing in a slightly different way: Pigeons act as if they weigh potential losses in food much more than wasted key pecks.

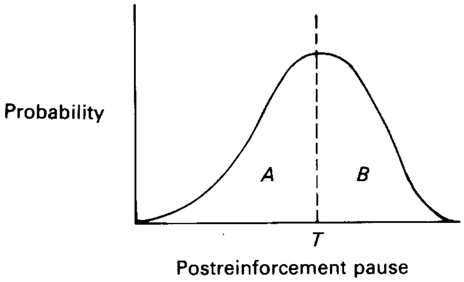


Figure 5.18. Theoretical pause (time-estimation) distribution of a pigeon on a fixed-interval *T* sec schedule.

SUMMARY

The effects of reward and punishment on behavior are obvious to all. Unaided by experimental psychologists, the human race managed long ago to discover that children and animals desist from punished behavior, and persist in rewarded behavior. Science has added to this rude knowledge in at least three ways. One is to define reward and punishment precisely. The result is the concept of a *reinforcement contingency.* Another has been to emphasize the role of time delays between response and reinforcer. And a third has been to design simple situations that allow us to explore the limits of animals' ability to detect reinforcement contingencies. We have begun to understand the intricate mechanisms that allow mammals and birds to detect subtle correlations among external stimuli (including time), their own behavior, and events of value.

In this chapter I have presented a brief history of operant and classical conditioning and described some common conditioning situations. I have explained the difference between *acquisition,* the process by which animals detect contingencies, and the *steady state,* the fixed pattern that finally emerges. The chapter discussed the concepts of *stability* and *equilibrium,* and how operant behavior, while usually stable, is occasionally unstable or metastable in response to changing conditions. I spent some time on *molar feedback functions,* because these seem to determine molar adjustments to many simple reinforcement schedules, and also shed light on motivational mechanisms, which are the topic of the next chapter.

NOTES

[1] A drive implies "push" rather than "pull"; it sounds more like a mechanistic than a functional (teleological) explanation. Perhaps this is why the notion of drive was so appealing to some early behaviorists. The concept of preference structure is clearly functional, however. A preference structure, like the decision rules for *Stentor* discussed in the last chapter, implies no particular mechanism. The distinction is more semantic than real, however. Drives were treated very much like goals by their advocates, so I do the concept no violence by treating drives as a primitive form of preference structure.

For good accounts of theories of primary and secondary drives (theories of which were developed most extensively by Hull and his followers) see Osgood (1953), Kimble (1961), and Bower and Hilgard (1981).

[2] *Recent history of reinforcement theory.* The theoretical foundations for Thorndike's work were laid by the English philosopher Herbert Spencer (1820-1903), from whom Darwin borrowed the phrase "survival of the fittest." Spencer applied to trial-and-error learning a version of the Darwinian theory of adaptation via variation and selection, which is essentially the position I am advancing here. Similar views were outlined by the British comparative psychologist Lloyd Morgan (1852-1936), who is perhaps best known for his application of the principle of parsimony to explanations of animal behavior: Never invoke a higher faculty if a lower one will do (Lloyd Morgan's canon).

American comparative psychology was distracted from these rational, biologically based doctrines by the strident tones of J. B. Watson's behaviorism, which was no more objective or parsimonious than Thorndike's views, but was easier to understand. Truth must be understood to be believed, but ease of understanding is a poor guide to validity. Nevertheless, Watsonian behaviorism captured the public imagination, and left American animal psychology with a taste for simplism from which it has never recovered. After his initial contributions, Thorndike moved away from animal psychology, and for some years the field was left largely to three men and their followers: Edwin Guthrie, E. C. Tolman, and Clark Hull. The first two were eclectic in spirit and neither founded a school. But Hull, at Yale, an intellectual descendant of Watson, organized his laboratory along Stakhanovite lines and produced a stream of energetic and dedicated disciples who colonized several other departments of psychology. Soon his students, most notably Kenneth Spence at Iowa, were producing disciples of their own and Hullian neobehaviorism, as it came to be known, became the dominant approach to experimental psychology.

In the mid-1930s B. F. Skinner at Harvard proposed an alternative to Hullian stimulus-response theory. In some ways almost as simplistic as Watson's views, Skinner's atheoretical approach had the advantage over Hull's far-from-elegant theory of a simple and effective new experimental method, the Skinner box. The social momentum of the Hullian movement ensured its continued dominance until the late 1950s, after which it was superseded by Skinner's operant-conditioning method. Skinner's philosophical approach to behavior also gained many adherents. Though currently out of fashion, adherents to one or other variant of radical behaviorism remain a vigorous vestige of the behavioristic tradition.

Hull's fall and Skinner's rise are probably attributable to growing impatience with the unwieldiness of Hull's proliferating system, and the collective sense of liberation felt by those who turned away from it to Skinner's new experimental method. Skinner wrote persuasively of the power of an "experimental analysis" of behavior and converts to *operant conditioning,* as the field came to be called, found exciting things to do exploring the novel world of reinforcement schedules. It took some time before others began to question the relevance to behavior-in-general of the elaborations of reinforcement-schedule research which, in its heyday, became a sort of latter-day experimental counterpoint to the Baroque fugue of Hullian axioms and corollaries.

Skinner's theoretical position is an intriguing one. On its face it is simple to the point of absurdity. It is a theory and an epistemology, yet claims to be neither. This ambiguous status has conferred on it a measure of invulnerability. Its basis is the principle of reinforcement: All operant behavior (there is another class, *respondent* behavior, that is excluded from this) is determined by the reinforcement contingent upon it. Taxed to identify the reinforcers for different kinds of behavior, Skinner responds that the principle of reinforcement is a definition, not an explanation, thus begging the question. Pressed for the reinforcement for behavior on which nothing external depends, recourse is often had to "self-reinforcement," which is self-contradictory. By giving the appearance of a theory, but disclaiming the responsibilities of one; by providing a method which promises to find a reinforcer for everything — and by invoking a mass of data on schedules of reinforcement exceeding in reliability and orderliness anything previously seen, Skinner was for many years able to overwhelm all argument and present to the world the vision of a powerful and consistent system that promised to remake human society.

Aristotle distinguished four kinds of causes, two of which are of contemporary interest: the efficient cause and the final cause. Efficient cause is *the* cause of mechanistic science, which considers no other an adequate explanation. The stimulus for a reflex response is a cause in this sense. Final causes have not been totally dispensed with, however. The economic concept of *utility,* which is another name for *subjective value,* is a final cause in the Aristotelian sense, because consumers are presumed to adjust their behavior so as to maximize their utility: The consumer's market basket is explained not by the causal factors underlying each purchasing decision, but by reference to the utility of the final bundle. It is obvious that reinforcement is a final cause in the Aristotelian sense, and plays the same role in Skinnerian theory as utility does in economic theory. The parallels are close. For Aristotle, everything has its final cause; for radical behaviorists (Skinner's term), every operant behavior has its reinforcer. In popular detective fiction, the first rule is to look for a motive, which usually involves a woman. Skinner will have nothing to do with motives, but his views can, nevertheless, be epitomized as *cherchez la renforcer.*

The intimate conceptual relation between utility and reinforcement theory has led recently to increasing application of economic ideas to behavioral psychology, and vice versa; more on this in Chapter 7.

In addition to the references given earlier, fuller discussions of behaviorism and the law of effect can be found in Boring (1957), Hearst (1979), Postman (1947), Wilcoxon (1969), the volumes edited by Koch (e.g., 1959), and a collection of theoretical reviews by Estes et al. (1954). See also Baum (1994) and Richelle (1995) for sympathetic accounts of Skinnerian behaviorism and Staddon (2001a) for a critical account.

3. The terminology in this area is confusing. The terms I have used are probably the simplest and most intuitively obvious. Skinner used the same terms in a slightly different sense, however, and many follow him. Skinner attended to the fact that a reinforcing effect may be produced either by presenting a stimulus (e.g., food) or by removing one (e.g., electric shock). The first case he called *positive reinforcement;* the second, *negative reinforcement.* The symmetry breaks down when applied to suppressive effects, however. Suppression of behavior by the presentation of a stimulus contingent on it is termed *punishment* — I have termed this *negative reinforcement;* but there is no simple term for the suppression of behavior by the response-contingent removal of positive stimulus.

The only really precise way to represent these various cases is by means of feedback functions, which I take up in connection with the concept of reinforcement contingency. In the meantime, the terminology I have adopted is probably less confusing than the alternatives because it focuses on the direction of change of the behavior, rather than the kind of stimulus change necessary to produce that change. For example, is *heat* a positive or a negative reinforcer in Skinner's sense? It is difficult to answer this without getting into unprofitable discussions about whether it is the presence of heat or the absence of cold that is reinforcing. The recognition that reinforcement is equivalent to feedback renders such disputes unnecessary.

4. The method of studying reward and punishment via its effects on an "arbitrary" response was independently invented by B. F. Skinner (1932) in the United States and G. C. Grindley (1932) in England. Grindley used the response of head-turning by restrained guinea pigs; Skinner used lever pressing by freely moving rats. Skinner's approach has dominated subsequent research, for at least two reasons: His method was obviously easier to implement than Grindley's and he discovered that food need not follow every response. This finding led to the concept of a *reinforcement schedule* and the extensive study of the properties of reinforcement schedules and their implications for the mechanisms of operant behavior.

5. I will use the term *reinforcement* in a purely descriptive sense, to refer to the operation of presenting a reinforcer to an animal. Unfortunately it is also often used as an explanation, based on theories (now largely abandoned) derived from the original law of effect that assumed that presenting a reinforcer contiguous with a response automatically "strengthened" (reinforced) that response or its connection with current stimuli.

6. See Staddon (1965) for a fuller discussion of this example.

7. Pavlov focused almost entirely on the salivary conditioned response. Recent experiments using the Estes-Skinner procedure tend to look just at the suppression measure. Other recent work has emphasized the earlier conclusion of Zener (1937) that these simple changes are perhaps the least important of the many effects of classical-conditioning procedures, however. I return to a discussion of these other effects in Chapter 13 in connection with the acquisition of behavior. The question of *why* a stimulus associated with shock should suppress food-reinforced operant behavior is also taken up later.

8. If the entries in the four cells of the contingency table are labeled *a, b, c,* and *d* (reading from left to right, top to bottom), then x^2 is given by

$\chi^{2} = (ad - bc)^{2} / (a+c)(c+d)(b+d)(a+c); N = a+b+c+d$. Other measures of contingency are information transmission and the d' measure of signal detection theory (see Luce, 1963, Green & Swets, 1966). See Gibbon, Berryman, and Thompson (1974) for an extensive (though sometimes hard to follow) discussion of the concept of contingency as applied to both classical and operant conditioning procedures.

9. A contingency table says nothing about the direction of causation; it is completely symmetrical. Consequently, we could equally well have divided the entries in Table 5.4 by the column, rather than row, totals, to obtain the conditional probabilities of *S* and ~*S* given *Sh* and ~*S,* rather than the reverse. Since the animal's interest is obviously in predicting *Sh* from *S,* rather than the reverse, the row totals are appropriate divisors.

10. φ here was computed using the time-based method (see Gibbon et al., 1974, for a fuller account).

11. The meaning of "negative contingency" is obvious when an exact feedback function can be derived. It is less clear when no analytic function is available. Perhaps the most general definition is in terms of the sign of the first derivative, $dR(x)/dx$. The sign of a contingency then refers to the effect on the direction of change of reinforcement rate of small changes in response rate. For complex schedules, the sign of $dR(x)/dx$ may not be the same across the whole range of x. This will always be the case for any schedule designed to restrict x to a limited set of values. For example, on spaced-responding schedules, a response is reinforced only if it follows the preceding response by t sec or more. Thus, when the animal is responding rapidly, so that the typical interresponse time (IRT) is less than t, the contingency is a negative one: Decreases in response rate bring increases in reinforcement rate. But when the animal is responding slowly (IRT> t), the contingency is positive: Increases in response rate bring increases in reinforcement rate.

12. This function was derived as follows. If responding is random in time, then the probability that an IRT will fall between zero and *T,* the postponement value, is given by

$$P(\langle T) = \int x \bullet \exp(-xt)dt, \tag{N5.1}$$

which equals

$$= 1 - \exp(-xT). \tag{N5.2}$$

The time taken up by IRTs in this range is given by

$$T_s = \int_0^T tx \cdot \exp(-xt)dt \tag{N5.3}$$

$$= 1/x - (T + 1/x)\exp(-xT).$$

Similarly, the probability that an IRT will be greater than *T,* hence fail to avoid shock, is

$$P(>T) = \exp(xT). \tag{N5.4}$$

The time taken up by IRTs in this range is then just

$$T_u = T\exp(-xT), \tag{N5.5}$$

because of the assumption that the animal responds immediately after shock. The shock rate, $R(x)$, is just the number of shocks received divided by the total time, $T_s + T_u$. From Equations N5.4, N5.3, and N5.5:

$$R(x) = x\exp(-xT)/(1 - \exp(-xT)),$$

which is Equation 5.7 in the text.

The variable x in these equations is not the actual response rate, but the corresponding parameter of the exponential distribution. When $x \gg 1/T$, x is approximately equal to x_{ac}, the actual response rate. Otherwise Equation 5.7 must be corrected by the relation

$$x_{ac^{\cdot}} = x/(l - \exp(-xT)), \tag{N5.6}$$

which is derived from the relation

$$x_{\mathrm{ac}} = l/(T_{\mathrm{s}} + T_{\mathrm{u}}).$$

13. For a review see Catania (1981) and several of the other articles in the book edited by Zeiler and Harzem.