16
LEARNING, III: EXPERIMENTAL ANALYSIS

One sign that the field of animal learning is still relatively immature, is that textbooks are still dominated by the study of particular experimental procedures, rather than the principles and processes that have been derived from them. This chapter discusses a number of these procedures from the point of the principles derived in earlier chapters.

Conditioned reinforcement

Schedule-induced behavior occurs when periods of food and no-food are signaled by time. What happens when periods with different payoff conditions are signaled by nontemporal stimuli, like lights, sounds or spatial locations? The simplest case is when stimulus changes are independent of responding; these are the *multiple schedules* discussed at length in Chapters 11 and 12. The allocation of behavior here parallels the fixed-time schedule: interim (facultative) activities predominate in the components associated with a low rate of reinforcement, the terminal (instrumental) response predominates in components associated with a high reinforcement rate.

More complex effects come into play when the stimulus changes depend on the instrumental response. There are two main effects to consider: stimulus change as an aid to *memory*, and stimuli as guides to behavioral *allocation*. Sometimes these two effects work together, and sometimes they conflict.

The memory effects are seen most clearly in the acquisition of a response. For example, suppose we attempt to train a pigeon to peck a white key for food reward, but delay the food for 5 seconds after each peck. Even if the animal does eventually peck the key, the effect of each reward is likely to be small, the animal will obtain little food, and training may be unsuccessful. The likely reason is that it is difficult for the animal to pick out the peck, a brief event preceded and followed by other activities, as the best predictor of food, unless peck and food occur close together in time.

We can make the bird's task much easier in the following way: We train him in two phases. In phase 1, the white response key occasionally turns green for 5 seconds, after which food is delivered. No response is required. This is of course an autoshaping procedure, and the pigeon will soon be pecking the green key, but this is not in itself important (the results will be the same even if we prevent the animal from pecking the key during this phase, either by restraining him, or by covering the key with a clear window). In the second phase, pecks on the white key are immediately followed by the green stimulus plus food after 5 seconds. The time relations between pecking and food are exactly the same in the second phase as in the first — yet the pigeons will rapidly learn to peck the white key if given 5-sec of green as a signal. It is not even necessary to begin with the autoshaping procedure: pigeons dumped directly into phase 2 will also learn. Why is this added-stimulus procedure more effective than the simple delay procedure?

The two-stimulus procedure in phase 2 is termed a *chained schedule*: peck → stimulus → food. The green stimulus is used as a reinforcer and seems to act like one. On the other hand, it gains its power not innately (or early in development), but by virtue of its pairing with food. Hence it is termed a *conditioned* or *secondary* reinforcer. The green-key conditioned reinforcer aids conditioning for two reasons: First, it bridges the temporal gap between the peck (the real cause of the food) and its consequence (the food). Rather than having to remember a brief event occurring 5 sec before its consequence, the animal has only to remember that pecking leads to stimulus change — since the peck→stimulus-change delay is negligible, this presents no difficulty. Second, the rate of food delivery in the presence of the green stimulus is much higher than

the average rate of food delivery in the apparatus. Hence, the green stimulus is a predictor of food and has value of its own, although the value is relatively transient and depends upon reliable delivery of the food. If food ceases to occur, the green stimulus will soon lose value, as choice tests have repeatedly demonstrated. The ubiquitous hill-climbing tendency will lead the pigeon to peck at the white key so as to produce the higher-valued green key.

Looking at the green key as a memory aid leads to different predictions than looking at it as a surrogate for food, a "reinforcer" that can "strengthen" behavior. Let's look at some of the problems with the "reinforcer" model:

If a one-link chained schedule is effective in maintaining key pecking, why not two, three, or *N* links? Two links are generally effective. For example, suppose we train a pigeon on the following sequence, where the notation S_i:peck$\rightarrow S_j$ indicates that a peck in the presence of stimulus S_i produces stimulus *j:*

$$S_1\text{:peck}\rightarrow S_2\text{:peck}\rightarrow S_3\rightarrow\text{food}\rightarrow S_1$$

which corresponds to a two-link chain. Providing the durations of each stimulus are appropriate (t_2 and t_3 should not be much longer than t_1; the ideal arrangement is for t_1 to be much longer than t_2 and t_3) pigeons will learn to peck S_1 to produce S_2, and peck S_2 to produce S_3. But the process cannot be continued indefinitely. Imagine a chained schedule of the following form:

$$S_1\text{:}t\text{>}T\text{, peck}\rightarrow S_2\text{:}t\text{>}T\text{, peck}\rightarrow\ldots S_N\text{:}t\text{>}T\text{, peck}\rightarrow\text{food}\rightarrow S_1\text{:}\ldots$$

The term S_i:t>T, peck$\rightarrow S_j$ denotes a fixed-interval T-sec schedule in the presence of S_i (a peck more than T sec after the onset of S_i is reinforced by the appearance of S_j). The whole sequence denotes a chained schedule totaling *N* fixed-interval links. How many such fixed-interval links can be strung together and still maintain responding in S_1, the stimulus most remote from food?

The answer is, not more than five or six. The top record in Figure 16.1 shows a cumulative record from a pigeon trained with six fixed-interval 15-s links (the recorder pen reset after food). Long pauses occur after food and the times between food delivery are always much longer than the 90 sec minimum prescribed by the schedule. An additional link would have caused the pigeon to stop responding altogether. With longer fixed intervals, five links are the upper limit.

The middle record in Figure 16.1 shows a typical performance when the stimulus change from one fixed-interval to the next is eliminated, but all else remains the same (this is termed a *tandem schedule*). There are two notable differences between performance on the tandem and chained schedules: the tandem schedule is not subject to the long postfood pauses on the chained schedule; and the terminal response rate (i.e., the slope of the cumulative record in the last fixed interval before food) is much higher in the chained schedule. These differences appear even more clearly in the bottom record, which shows both types of schedule successively in the same pigeon.

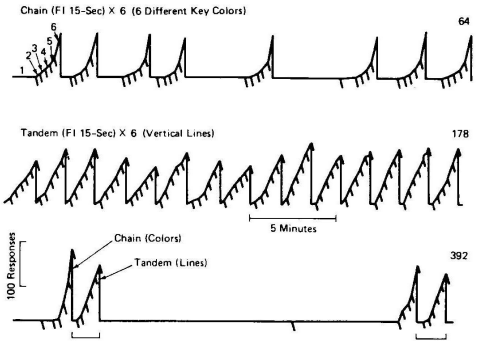

Figure 16.1. *Top***:** Cumulative record from a pigeon well trained on a 5-link (6-stimulus) chained fixed-interval 15-sec schedule (the recorder pen reset after food, and blips indicate stimulus changes). Center: Record of a pigeon trained on the tandem schedule equivalent to the 6 stimulus chain (a single stimulus, vertical lines, was on the key throughout, but the six successive fixed-interval schedules were in effect between food deliveries). *Bottom*: Record of a pigeon exposed to both the chained and tandem procedures in alternation. (From Catania, 1979, Figure 8-12.)

Obviously one problem with the simple conditioned-reinforcement idea is that it cannot explain results like those in Figure 16.1: Why should five or six links be the limit? Why should the pigeons respond *less* in the first link than if there were no stimulus change? These problems are usually handled by adding another process to conditioned reinforcement. Thus, two things are usually thought responsible for the differences between chained and tandem schedules illustrated in Figure 16.1: (a) The relative proximity to reinforcement (food) of each stimulus in the series. (b) Conditioned reinforcement, that is, the contiguity between pecking and the transition to a stimulus closer to food.

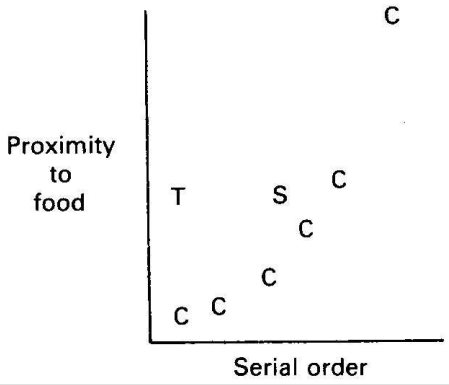

Figure 16.2. Stimuli arranged according to their serial order and proximity to food in three procedures: chained schedules (C's), tandem schedule (T), and scrambled chain schedule (S).

Proximity to food clearly exerts the major effect. It is illustrated in Figure 16.2, which plots the serial order of each stimulus against its temporal proximity to food. The *C*'s (for "chain") in the figure denote stimuli in the chained schedule, where serial order is perfectly correlated with proximity to food. The "*T*" denotes the serial order and average proximity to food of the single tandem stimulus. Clearly, response rate in the presence of each stimulus is directly related to its relative proximity to food: rate in the chained procedure is higher, for the terminal link, and lower, in the initial link, than the average rate in the tandem schedule. This conclusion is confirmed by the *S* in the figure, which shows the effect of scrambling the order of the chained stimuli, so that each one appears equally often in each serial position. Pigeons so trained respond at essentially the same rate in the presence of each stimulus, as suggested by the single average value for both serial order and proximity to food.

Time by itself seems to have a relatively small effect on the tandem schedule, as indicated by the gradual increase in response rate with postfood time.

The concept of conditioned reinforcement (that is, the response contingency between pecking and stimulus change) adds little to our understanding of chained schedules. Experiments have shown that the contingency makes surprising little difference to performance (see Catania, Yohalem & Silverman, 1980, and review in Staddon & Cerutti, 2003). Providing the response contingency for food in the terminal link is maintained, it can be omitted in earlier links with little effect on key pecking, as long as stimulus changes continue to take place as before (i.e., the fixed-interval contingency is replaced with a fixed-*time* contingency — this is termed a fixed-interval *clock*, because the successive stimulus changes signal the lapse of postfood time). Behavior on chained schedules is determined by temporal proximity to food in the same way as behavior on multiple schedules.

The pattern of terminal responding on chained schedules is complementary to the pattern of interim activities, which occur preferentially in components whose reinforcement rate is below the average. The increasing pattern of terminal responding, and the tendency to long pauses in the first link can be derived by the same argument used to explain behavioral contrast in Chapter 11. For example, the immediate postfood period (i.e., the first link) predicts the absence of food, so that interim activities will tend to occur preferentially at that time. Increases in interim activities in the first link reduce the probability of a terminal response. Since a response is necessary for the transition to the second link, this tends to lengthen the first link. The longer the first link, the lower its proximity to reinforcement relative to others, which further reduces the probability of a terminal response, further lengthening that link, and so on. Moreover, the more time devoted to interim activities in the first link, the less likely they are to occur in later links (because of the satiation-deprivation processes described in Chapter 11), allowing more time for

the terminal response. The whole process, then, favors the kind of temporal gradient which is actually observed, in which terminal response rate increases through the chain, winding up highest in the last link.[1]

Conditioned reinforcement in two-link chained schedules facilitates the acquisition and maintenance of behavior because it aids memory. A related procedure, *second-order schedules*, has its effects by impairing memory. The basic phenomenon is illustrated in Figure 16.3. The top cumulative record shows the relatively high response rate and scalloped pattern produced in a well-trained pigeon by splitting up a fixed-interval 60-min schedule into fixed-interval 4-min components, each terminated by a response-contingent brief (0.5-sec) stimulus. Food follows (i.e., is paired with) the last such brief stimulus in the 15-component cycle. The bottom record shows the low, unpatterned responding generated by the comparable tandem schedule. The tandem performance shows that without the brief stimuli, response rate is very low; and in the absence of time markers, there can be no 4-min scallops.

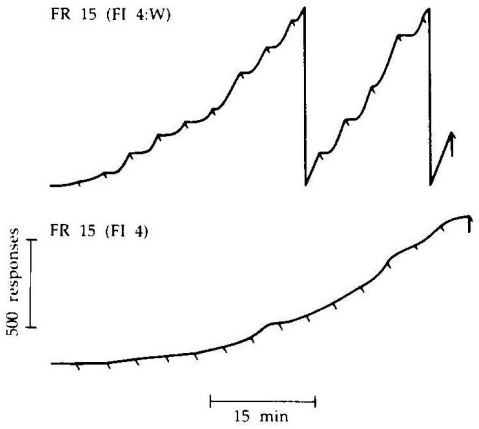

Figure 16.3. *Top record*: cumulative record of the performance of a single pigeon on a second-order FI 60-min (FI 4-min) schedule. Every 4 min, a peck produced a brief .7-sec stimulus on the response key (diagonal blips on the record; this is the FI 4 component); at the end of the 15th such presentation, food was also delivered (the record resets; FI 60). Thus, the brief stimulus was paired with food once an hour – and also signaled a food opportunity after 4 min. *Bottom record*: performance on the same schedule, but with no brief-stimulus presentations (tandem schedule); blips show FI-4 components. (After Kelleher, 1966.)

The brief stimuli seem to act by interfering with the animal's recall for the most recent food delivery. In Chapter 13, I showed how the scalloped pattern on fixed-interval schedules depends upon the animal's ability to recall the most recent time marker. The best temporal predictor of food in the second-order procedure shown in Figure 16.3 is food — but it is temporally remote (the interval is 60 min long) and the time is filled with periodic brief stimuli, the last-but-one of which is also a temporal predictor of food: food is always preceded by a brief stimulus 4 min earlier. The greater validity of food ($p(F|t>60) = 1$) seems to be outweighed by the greater frequency and closer proximity to food of the less-valid ($p(F|t>4) = 1/16$) brief stimulus — so that the brief stimulus seems to overshadow food as the effective time marker. Since the animal does not know where it is in the interval (in confirmation of this, the temporal gradient after the first postfood brief stimulus in second-order schedules is much less than in comparable fixed-interval schedules), and anticipates food every 4 minutes, response rate is naturally higher in the second-order than the tandem schedule. As this argument leads one to expect, brief stimuli have their largest effects with long interfood intervals.

The effects of brief stimuli were at one time attributed to the pairing of the final stimulus with food. Subsequent work has devalued the importance of pairing, which seems now to be important mainly for the initial acquisition of the pattern (see Squires, Norborg, & Fantino, 1975; Stubbs, 1971; and Staddon, 1974).

As with *primary* reinforcers such as food or electric shock, conditioned reinforcers seem to exert their major effect through inducing factors — the Pavlovian relations between particular synchronous or trace stimuli and the frequency of primary reinforcement. The response contingency plays a role in two ways. It may aid acquisition of the correct response and correct any tendency for the response to drift away from the effective form; and through the feedback function it maintains the stimulus-reinforcer relations that sustain the response.

Conditioned emotional response

Many, perhaps most, contemporary experiments on classical conditioning use a mixed procedure in which the effects of a CS are assessed indirectly. Rather than measuring the direct effect of a CS on an autonomic response such as salivation or pupillary dilation, it has turned out to be more convenient to study its indirect effects on food-reinforced lever pressing. Rats are first trained to press a lever for food, delivered on a variable-interval (VI) schedule. As we have seen (Chapter 5), after a little training this schedule maintains a moderate, steady rate of response — an ideal baseline for the study of other variables. Stimuli of perhaps 60-s duration are occasionally superimposed on this VI baseline. Rats soon habituate to the added stimulus, and response rate is the same in the presence of the superimposed stimulus as in the background stimulus before and after. In the final phase, electric shock is delivered in the presence of the superimposed stimulus (CS). Within a few stimulus-shock pairings, the onset of the CS causes a reduction in the rate of lever pressing, which has been termed *conditioned suppression* or the *conditioned emotional response* (CER) — or the Estes-Skinner procedure, after its inventors (1941). The amount of suppression in the CS, relative to the just-preceding baseline period, then provides a measure of the amount or strength of conditioning.

On the face of it, this suppression makes little adaptive sense. The rat has ample opportunity to learn that it cannot avoid or escape from the shock. By reducing its lever-press rate, the animal loses needed food reinforcers. What causes this apparently maladaptive suppression?

By now, the answer should be clear. The superimposed stimulus is a CS for shock. Once the animal has learned this relation, therefore, we may expect to see candidate defense reactions induced by the CS, according to the Pavlovian inference mechanisms I have discussed in this and the preceding chapter. These reactions will generally interfere with lever pressing, hence show up as a depression in lever pressing — the CER. These reactions persist even though they have no effect on shock for the same reason that key pecking persists under an omission contingency. The priors associated with a stimulus that predicts imminent shock simply outweigh the opposing effects of the food contingency for lever pressing.

This argument, and data of the Brelands discussed earlier, suggests that similar suppression effects might be produced even by a food stimulus on a food-reinforced baseline. For example, a feeder light that comes on 5 sec before the feeder operates will soon cause a hungry rat to approach the feeder as soon as the light comes on. If such stimulus-food pairings are superimposed on a VI 60-sec baseline, no one would be startled to observe a reduction in lever pressing during the stimulus, and this is indeed what occurs. Conversely, if the light were on the response lever rather than the feeder, lever pressing might well *increase* during the CS — because the rat now approaches the lever (and presses it) rather than approaching something incompatible with lever pressing.

Both these effects of a food CS have been widely observed.[2] As these examples suggest, the magnitude and direction of the effect depends on the duration of the CS (both absolute, and relative to the time between CS presentations), the type (e.g., tone, light, localizable vs. unlocalizable) and location of the CS relative to the feeder and the lever, and the magnitude and frequency of food in the presence of the CS. The general rule is that if the food rate in the presence of the stimulus is significantly higher than the rate in its absence (i.e., on the baseline VI schedule), the stimulus will tend to induce food-related behavior. The effect on lever pressing then depends on the nature of the induced behavior: if it is physically compatible with lever pressing, the effect of the CS is to facilitate lever pressing, if not, it is to suppress lever pressing.

Avoidance and escape

Electric shock has a number of paradoxical effects, which all derive from its strong inducing effects. Shock, and stimuli that signal shock, produces very stereotyped reactions from

most animals, and immobility ("freezing") is often dominant, especially if the shock is severe. For example, if a rat is presented with a train of brief electric shocks that can be turned off by some predesignated response, he may only learn the necessary escape response if it is part of the induced defense reactions. If the required response is to press a lever, it will be learned much more easily if lever *holding* is sufficient than if active *depression and release* is required. Holding is compatible with the induced reaction of "freezing" on the lever, whereas lever pressing requires that the animal periodically release the lever so that it can be depressed again. The inducing effects of shock make shock-motivated operant conditioning difficult for rats and pigeons for reasons not directly related to the supposedly indirect nature of avoidance schedules.

In all shock experiments, the animal's highest priority is to escape from the apparatus entirely. This is precisely what one might expect not only from common sense, but from the principles of positive reinforcement. In a food situation, animals detect the stimulus most predictive of food and approach it. In an aversive situation, they detect the stimulus or situation most predictive of the *absence* of shock, and approach that — this rule takes them at once out of the experimental apparatus, which is the only place they normally experience shock. Since escape is always prevented, anything they do in the experimental situation is in a sense second best, and not an ideal measure of the effect of the shock schedule. Studies of positive reinforcement don't suffer from this problem: a hungry animal is happy to be in a box where he gets food.

So-called "one-way" shuttlebox avoidance shows the importance of withdrawal. Animals shocked on one side of a long box, will immediately run to the other side when placed in the box. In contrast, "two-way" avoidance, in which the animal must run back and forth from one side of the shuttlebox to the other whenever a signal sounds, is harder for animals to learn. There is no safe place (within the shuttlebox), the animal's dominant response (escape) is blocked, and so learning is difficult.

Free-operant, nonlocomotor, avoidance procedures are of two main types: *discriminated avoidance,* and *shock-postponement* (Sidman-avoidance) schedules. In discriminated avoidance, a stimulus is occasionally presented; shock occurs after a few seconds, and then shock and stimulus both terminate (Figure 16.4, top). Granted the difficulty in ensuring that lever pressing is part of the animal's candidate set, in other respects the explanation for discriminated avoidance is the same as the explanation for conditioned reinforcement: The animal is presented with two situations, CS and CS-absence ($\overline{CS}$), one higher valued than the other. A response in the CS (the lower-valued situation) produces $\overline{CS}$, the higher-valued one. The response-contingent transition from CS to $\overline{CS}$ ensures that the animal can remember the effective response, and the higher value of (lower shock rate in) $\overline{CS}$ ensures that once the response occurs, it will be maintained. Experiments with shock-postponement schedules have shown that the pairing of the CS with shock is much less important than its effect in making the avoidance response easy to remember. Thus the improvement in shock rate (detected we know not how) is the factor that maintains avoidance responding.

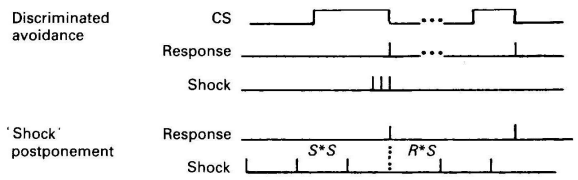

Figure 16.4. Shock-avoidance procedures. *Top*: discriminated avoidance - a response during the CS turns it off (avoidance); shock occurs at the end of the CS until a response occurs (escape). *Bottom*: shock postponement. Shock occurs at fixed intervals (the shock-shock – S*S – interval) unless a response occurs; each response postpones the next shock for a fixed interval (the response-shock – R*S – interval).

Shock postponement is illustrated in Figure 16.4 (bottom): brief shocks occur at fixed time intervals of say 10 s (this is termed the shock-shock or S*S interval); if the animal responds, usually by pressing a lever, the next shock (only the next) is postponed for a fixed time, say 15 s

(the response-shock or R*S interval), which may be the same as or different from the S*S interval. This is *fixed-interval shock-postponement*. No matter what the value of the R*S interval relative to the S*S interval, it pays the rat to respond at least once every t seconds, where t is less than the R*S interval. By so doing, it can avoid all shocks. Another version, variable- (or random-) interval shock-postponement, is very similar. If the rat does not respond, it receives random, brief shocks at a certain rate, say two per minute, defining an average intershock interval t_S. If it does respond, the *next* shock occurs at an average time t_R after the response, where t_R is greater than t_S.

Rats learn to respond on both procedures, albeit with some difficulty. Judged by the proportion of animals that completely fail to learn, these procedures are more difficult even than the two-way shuttlebox, and much more difficult than one-way shuttle avoidance, or avoidance where the response is running in a wheel. The source of the difficulty seems to be that lever pressing is often not one of the candidate responses made by animals in these situations. Weaker shock often aids learning, presumably by diminishing the tendency for strong shock to induce rigidly stereotyped behavior. Both shock-postponement procedures provide an opportunity for response selection by relative contiguity: of all activities, the effective response will on the average be the one most remote from shock. If proximity to shock tends to exclude an activity from the candidate set, then the animal should eventually arrive at the correct response as the one *least contiguous* with shock — and this response will generally be effective in reducing overall shock rate.

Relative contiguity is of course the process by which *punishment* (response-contingent negative reinforcement) selectively eliminates the punished activity.

Rats will sometimes respond on a shock-postponement schedule that has no effect on shock rate, providing each response produces a brief shock-free period. For example, Hineline[3] trained rats in an extensive series of experiments based on a 60-s cycle. A response lever was present at the beginning of each cycle. If no lever press occurred, the lever retracted after 10 s, and a single brief shock occurred at the end of the 11th second. The lever reappeared after 60 sec and the cycle resumed. If a lever press occurred, the lever at once retracted, the 11-s shock was omitted, but a shock occurred at second 39. Thus, a lever press had no effect on the overall shock rate, which was always 1 per 60 sec, but always produced a shock-free period of 39-t sec after lever retraction (vs. 1 s if no response occurred), where t is the time in the cycle when a response occurred. In later experiments, Hineline was able to train some rats to respond even if responding actually *increased* shock rate (several shocks, rather than just one, occurred after second 39). Control conditions, in which shocks occurred independently of responding, confirmed that the shock-free postresponse period was critical to maintenance of lever pressing in these animals.

It is hard to know how to interpret these studies. Not every rat shows the effects, and to explain response selection by the production of a safety period depends upon the rat representing the situation in a certain way. If the animal times everything from lever retraction (a very salient stimulus, because of the loud sound of the solenoid mechanism), then a lever press does indeed delay shock. On the other hand, if the animal assesses intershock time in some way, responding has no effect. Although shock never occurs when the lever is present, it does occur within one second of lever retraction, so that the lever does predict shock if the animal fails to respond. A one-second delay after a 10-sec stimulus is very short and perhaps small enough for the animal to treat the lever as a trace CS, so that the experiment is really a case of discriminated avoidance, rather than shock-postponement. A modified procedure in which the lever is present throughout each cycle, but all else remains the same, would almost certainly fail to sustain lever pressing.

What may be required as an adequate test of the delay-reduction idea is a modification of the variable-interval shock-postponement procedure in which a response produces a transient shock-rate reduction after variable amounts of delay. Instead of selecting the *first* shock after a

response from a distribution with a lower mean rate, one might select the second, third, or some later shock in this way. This technique encounters all the familiar problems of delay-of-reinforcement studies and might be difficult to design cleanly.[4] In this way one might perhaps trace out the negative delay-of-reinforcement gradient.

The conclusion of this discussion of avoidance is rather unsatisfactory. In a general sense, this behavior seems to follow the same rules as positive reinforcement: Shock is an hedonic stimulus and arouses the animal — although passive behavior ("freezing") often dominates unless special steps are taken to prevent it. The commonest active behaviors are attempts to withdraw from the source of shock or escape from the experimental situation. Since these tendencies are invariably blocked, the set of remaining candidates may be small. From this set, effective members are selected in ways not fully understood. The usual result is reduction of shock rate, but this cannot be used as a mechanistic explanation since we know little about how shock rate is assessed by the animal: The temporal relations between response and shock play some role, but it is clearly a rather weak one under many conditions. And some rats, under some conditions, can be trained to respond even if overall shock rate is thereby increased, so that shock-rate reduction is not acceptable as a genera; optimality account for avoidance behavior.

The small set of candidate activities induced by shock implies stereotyped behavior; as we will see shortly, this stereotypy can often be highly maladaptive.[5]

Set, response-produced shock and "learned helplessness"

Two necessary features of learning can produce maladaptive behavior under some conditions: (a) Learning depends upon *surprise*, and (b) Long training in situations with strong reinforcers (a very hungry animal and large or frequent food portions, strong electric shock) greatly reduces the candidate set. The stereotyped and inflexible behavior so produced has traditionally been known as behavioral *set*. These two characteristics can allow changes in the feedback function to go undetected, resulting in unnecessary, or even counterproductive, behavior.

Effective avoidance performance on a shock-postponement schedule means that the animal gets very few shocks. The few shocks that are received are important, however. Even well-trained rats show a reliable "warm-up" effect in each session, responding slowly or not at all at first. The few shocks received at this time induce avoidance responding so that later in the session almost all shocks are avoided. Suppose we take such an animal and turn off the shock generator, except for a few response-independent shocks at the beginning of each session. The animal now has almost no information to tell him that the world has changed and he need no longer respond: responding fails to occur at first, so he cannot learn it is ineffective; and never fails to occur later, so he cannot learn it is unnecessary. There is nothing to produce surprise, the first requirement for learning. The animal's only protection against this bind is to *sample* his environment either by responding early, or by occasionally *not* responding late, in a session. But sampling is just the overt manifestation of a large candidate set, and we have already established that schedules of severe shock, and protracted training, greatly reduce the candidate set. The animal does not sample, and if for some extraneous reason he fails to respond for a while later in the session, the absence of shock goes unnoticed. Avoidance responding persists almost indefinitely under these conditions.

Behavior maintained by the *production* of electric shock provides the most striking example of the effects of shock schedules in minimizing behavioral variation. An experiment by McKearney (1969; reviewed by Morse & Kelleher, 1977) is typical. Squirrel monkeys restrained in a chair were first trained to lever press on a schedule of shock postponement (S^*S = 10 sec, R^*S = 30 sec). When the typical steady rate had developed, a 10-min fixed-interval (FI) schedule of shock *production* was superimposed. Eventually, the shock-postponement schedule was phased out. The monkeys nevertheless continued to respond on the fixed-interval shock-production schedule and behavior soon became organized with respect to it: lever pressing fol-

lowed the typical scalloped pattern, increasing up until the moment of shock. As with food reinforcement, the shorter the fixed-interval duration, the higher the response rate. When shock was omitted entirely, the behavior soon ceased.

Rats in the typical Skinner box are less prone to the kind of rigidity shown by monkeys and cats in these experiments. But rats can be trained to produce electric shocks in a shuttle-box apparatus. Similar behavior has been established in humans and even goldfish. So-called self-punitive behavior is not an isolated phenomenon.

In McKearney's experiment, long training, physical restraint, and the highly aversive electric shock, all act to reduce behavioral variation. In addition, the transition from simple shock postponement to shock postponement plus FI shock is barely perceptible. The monkeys occasionally failed to avoid shocks on the postponement schedule, so an additional shock every 10 min made little difference. Moreover, even if the shock were detected, the animal could not cease responding without receiving many more shocks through the shock postponement contingency. All thus conspires to maintain responding in the face of occasional response-contingent shocks.

When the shock-postponement schedule was eliminated, leaving only response-produced FI shock, the sustaining factor for action, occasional shock, continued to occur. Moreover, given the relative effectiveness of avoidance behavior in well-trained animals, the omission of the shock-postponement schedule must have been imperceptible: many, perhaps most, of the shocks received were from the FI contingency before, and all were after, the change. Given a sufficient reservoir of behavioral variation, the monkey might have detected that almost any new behavior would lead to less shock than lever pressing. The monkeys' failure to do so suggests that the previous training had so reduced the candidate set that the necessary variation was lacking, so that responding continued indefinitely.

In some response-produced shock experiments, the process of systematically eliminating all candidates but the desired response is made quite explicit by a "saver" provision: For example, if the animal fails to respond at the end of a fixed interval, after a brief delay, several response-independent shocks may be delivered. This feature may later be phased out, but while it acts the animal is confronted with an avoidance schedule where responding makes perfect sense.

Behavior maintained by response-produced shock is not masochistic; the shock does not become pleasurable or positively reinforcing. Squirrel monkeys are transported to the experimental apparatus in restraining chairs partly because they would not otherwise choose to stay. The behavior is a misfiring of mechanisms that normally lead to escape and avoidance.

Shock in these studies seems to play two roles: (a) It provides a discriminative stimulus. The major discriminative property is to define the situation (see Chapters 10 and 14). Its most important characteristic for this purpose is its aversiveness, as we see in a moment. When shock is periodic (as in McKearney's experiment) it also provides a time marker. (b) Shock also motivates the behavior. Even highly effective avoidance behavior eventually extinguishes when shock is entirely omitted, but behavior maintained by response-produced shock persists indefinitely.

I have been arguing that all learning is selection from among a set of stimulus and response candidates, and we have seen that surprising results can come from procedures that seem to severely reduce the size of the candidate set. What if the set of response candidates is reduced to zero? The requirements for such a reduction should now be obvious: very severe, response-independent shock. Severity ensures that the candidate set will be small; response-independence that it will be empty, or at least contain only the "behavior of last resort" (usually freezing). These two characteristics define the phenomenon known as *learned helplessness*. In the original experiment (Seligman & Maier, 1967; for reviews see Maier & Jackson, 1979, Alloy & Seligman, 1979; and Glazer & Weiss, 1976) dogs were first restrained in a harness and given a series of severe, inescapable shocks. The next day, they were placed in a simple discriminated avoid-

ance situation: in each trial, when a CS came on, shock followed after ten seconds unless the dogs jumped over a low barrier. If they failed to jump, the CS remained on and shocks occurred for 50 sec. Thus the animals had an opportunity either to avoid, or escape from, the shock by jumping the barrier.

Normal dogs have no difficulty learning first to escape from shock, and then to avoid it by jumping as soon as they hear the CS. But the dogs pretrained with inescapable shock almost invariably failed to jump at all. Similar effects have been shown with a variety of species and aversive stimuli in appropriately engineered situations. The effects often generalize from one highly aversive stimulus, such as water immersion (very unpleasant for rats) to another, such as shock — emphasizing that the aversive property of the situation is the defining one for most animals.

Learned-helplessness is a dramatic example of the impairment of learning by *US preexposure*, discussed earlier: the dogs that received shock in the absence of the CS subsequently found it more difficult to detect the predictive properties of the CS. The magnitude of the impairment is surprising, but the severity of the shock can perhaps account for it.

The result also follows from the candidate-selection idea: pretraining with inescapable shock reduces the candidate set; subsequent training in a situation that permits avoidance or escape from shock is perceived as essentially the same (aversiveness defines the situation), so inaction persists. Shock continues to occur because the animal fails to avoid, so the process is stable — a self-fulfilling prophecy.

Some have argued that learned helplessness goes beyond the effect of CS preexposure. For example, a number of experiments have shown that the damaging effects of inescapable shock can be mitigated if animals have some prior experience with escapable shock. But the inferential explanation for US preexposure and related effects given earlier implies that the animal's candidate set should be determined by *all* its exposures to shock, not just the most recent, so this result does not justify placing learned helplessness in a special category.[6]

The effects of prior experience on learned helplessness are just what one would expect from the principles of memory: Animals exposed only to response-independent shock, looking back will see no effective response; animals with some experience of escaping shock see some possibilities for active escape that should transfer to the new shock situation. The importance of memory is also emphasized by the time limitations on the transfer from the response-independent shock to the test situation: if the two are separated by more than 24 hours, the dogs succeed in learning the avoidance response. As the memory of response-independent shock fades, the situation becomes more novel, the candidate set is in consequence larger, and the animal learns.

Learned-helplessness and response-produced-shock behavior can both be abolished by increasing behavioral variation. For example, helplessness can be overcome by physically helping the dog over the barrier — forcing it to sample. We have already seen that prior training with escapable shock leads via transfer to a reservoir of behavior that prevents helplessness. Other experiments have shown that changes in shock intensity and other features of the situation may mitigate helplessness — presumably by reducing transfer from the inescapable-shock situation. All generalization decrement (of which this is an example) represents increase in behavioral variation.

As this analysis leads one to expect, these maladaptive behaviors are metastable, that is, if behavior is disrupted in some way, it may not return to its previous form. For example, if responding that produces electric shock is somehow prevented, when the block is removed, the behavior is likely not to recover.

Learning is a way for animals to use their limited resources for action and computation as efficiently as possible. Particular activities are more likely to occur in situations where they will be of use — food-related activities when food is likely, defense reactions where there is

threat, and so on. In this way, energy is conserved and risk minimized. *Sampling*, variation in behavior so as to discover the properties of a situation, and *exploitation*, making use of what is learned, are necessarily in opposition. Without sampling, nothing can be learned; but without exploitation, there is no benefit to learning. Shock-produced behavior and learned helplessness are simply illustrations of one of the two kinds of mistake an animal can make during the process of learning: It can sample too much, and waste resources, or it can sample too little, and miss opportunities. Lacking omniscience, there is no way for an animal, or a man, to be certain of avoiding these two errors. The very best guess a creature can make is that when the cost of error is severe, and he has developed some way of coping, he should stick with it: don't mess with a winning (or at least, not-losing) combination. In human social life, the more serious the decision, the more it is surrounded with form and ritual: from cockpit drills to marriage — when the outcome is momentous, a rigid pattern of coping tends to emerge.

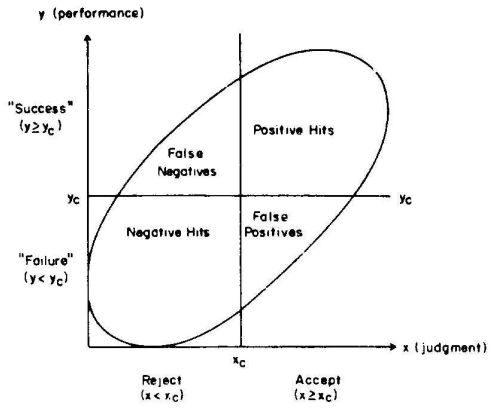

Figure 16.5. The relation between selection and performance measures and criteria for selection and performance, given an imperfect selection measure. (After Einhorn & Hogarth, 1978.)

This proneness to fixity in critical situations reflects not stupidity, but intelligence. Simple animals don't show these rigidities because their representation of the world is so primitive they cannot afford to rely on it. In place of rigid patterns, tied to well-specified situations, they must waste resources in random searching — which protects them from traps, but limits their ability to allocate their resources efficiently (see Chapter 3). No one has shown learned helplessness in a pigeon, and rats only show it in situations where even unshocked animals learn slowly. It is easy to show in people, and terrifying shock is not needed. The phenomenon is well known to sociologists under the name of *self-fulfilling prophecy*.

Recent work in human decision making has amply documented how people limit their sampling in ways that can lead to poor decisions. Consider, for example, how competitive research grants are awarded, how students are admitted to a selective college, or how personnel are selected. In every case, the thing on which a decision is based (the *selection measure*) is necessarily different from the final behavior that is the goal of the selection (the *performance measure*): research performance for the research grant, success in college and in later life for college admission, or on-the-job performance for personnel selection. Since performance cannot be measured at the time of selection, some indirect measure must be used, such as evaluation of a research proposal by a peer group (for a research award) or some combination of test scores and scholastic record (for college admission). Invariably a selection measure is chosen that will allow applicants to be ranked. This procedure assumes that the measure is positively related to the criterion, that is, people with high measure scores are likely also to score highly in terms of the performance measure: highly ranked students should do well in college, highly ranked research proposals should lead to high-quality research.

The decision situation is illustrated in Figure 16.5. The horizontal axis shows the selection measure (test score, for example); the vertical axis shows the performance measure (grade-point average, say). The performance-and selection-criterion values are shown by the crossed lines. Each individual can be represented by a point in this space whose coordinates are his scores on the test and in reality (performance score). The oval area shows the region where most individuals will lie, assuming that the test has some validity, but is not perfect (all points would lie on a straight line for a perfect measure).

The decision-maker obviously has two tasks here: to find the best rule for picking people, and to pick the best people. These two tasks are in opposition; they are just the dichotomy between sampling and exploitation we have already discussed. To pick the best people, the decision-maker need only set his selection criterion as high (as far to the right) as possible. In this way his proportion of *positive hits* (people above the *performance* criterion) will be as high as possible. But to estimate the validity of his selection measure, he also needs to sample individuals to the left of the criterion in Figure 16.12, that is, he needs an estimate of the proportion of false negatives — which means admitting some students with poor scores, or awarding research grants for poorly rated proposals. The necessary sampling of people to the left of the selection criterion almost never occurs,[7] partly because of the cost of false positives to the decision-maker (bad things happen to the personnel manager who hires a loser, but a winner missed costs him little), but partly also because many decision makers are unaware of the need to keep up to date on the validity of their decision rules.

People are just as dumb as animals in their unawareness of how poorly they are sampling. For example, in one famous experiment using the *Wason selection task* (see review in Cosmides & Tooby, 1992), subjects were presented with four cards lying on a table. A single letter or number (a, b, 2 or 3) was visible on each card. They were then told to check the truth of the statement "All cards with a vowel on one side have an even number on the other." To check positive instances, the card with "a" should be turned over; all subjects did this. But to check negative instances, the appropriate choice is the card with "3" visible; none turned over the "3". Many chose the "2" card, which of course adds nothing to the choice of the "a" card.

Both animals and people show this hill-climbing bias in favor of positive hits in experiments designed to test the so-called "information hypothesis" for *observing behavior*. In an observing-behavior experiment, subjects are offered the opportunity to produce a stimulus that tells them whether a reinforcer is likely or not, but has no effect on the actual availability of the reinforcer. For example, suppose that food for hungry pigeons is scheduled on a variable interval 60-s (VI 60) schedule for pecking on the left key, which is normally white. Pecks on the right key have no effect on food delivery, but turn the left key green if the VI is due to make food available within 30 seconds. Under favorable conditions, pigeons will soon learn to peck the right key because the rate of food delivery in the presence of green is higher than its rate in the situation as a whole, the standard conditioned-reinforcement result. And they peck for green on the right even though the observing response has no effect on the overall rate of reinforcement. A procedure that can give the animals essentially the same information is one where the left key is normally white, as before, but a peck on the right, "observing" key turns it green if food is *not* to become available in the next 30 sec. Pigeons will not peck the "observing" key under these conditions because it is associated with a rate of reinforcement lower than the overall average. This preference for "good news" is the same as the preference for positive hits in the human experiments. It reflects the universality of the hill-climbing heuristic (follow the direction of improving outcomes) discussed in connection with choice in Chapter 9.

Extinction

Extinction is both the abolition of an existing reinforcement schedule (procedural definition), and the decrease in the previously reinforced response that usually follows (behavioral definition). It involves the same learning mechanisms as the original acquisition of the response. The major difference is in the repertoire with which the animal confronts the changed situation, and the feedback function. In acquisition, the initial repertoire is variable and exploratory, appropriate to a novel situation. At the beginning of extinction, a single response class typically dominates. In acquisition, the animal's task is to detect a positive contingency between behavior and reinforcer; in extinction to detect the absence of any effective response. The task in extinction is more difficult, in the sense that in acquisition there is a "right answer," whereas in extinc-

tion there is none. Some say that "extinction is slower than conditioning," but there is no real evidence for this: the tasks are different, not the underlying processes.

Extinction is rarely immediate. For example, given a pigeon well-trained to peck on a variable-interval schedule, the cumulative record of responding on the first day of extinction will follow a negatively accelerated pattern of slow decline, with a complete cessation of responding only after several hours. When the animal is returned to the apparatus the next day, however, responding will resume at almost its old pace. This is termed *spontaneous recovery*. Spontaneous recovery makes great sense from a functional point of view — after all, things may have changed after a delay, and perhaps the mass of previous successful experience before the single extinction day is in some sense worth more than that one negative experience. The resemblance to habituation, and recovery from habituation after lapse of time, is also striking.

In terms of mechanism, spontaneous recovery is most parsimoniously explained as a reflection of well-established memory processes, rather than by making reference to the hypothetical stimulus difference between the beginning and end of an experimental session. Looked at in these terms, spontaneous recovery is a consequence of Jost's Law (see Chapter 13), the gain in influence of old experiences at the expense of more recent ones with the passage of time. Thus, at the end of the first extinction session, the recent experience of extinction is decisive, and the animal ceases to respond. But at the beginning of the next session, the experience of reinforcement in numerous past sessions reasserts its influence, and responding resumes. Soon, the growing period without reinforcement again exerts its effect and responding ceases. The same process is repeated, with diminished effect, the next day. As a backlog of days without reinforcement is accumulated, these experiences exert a growing effect at the beginning of each session, and behavior is finally abolished.

The memory property described by Jost's law is also involved in sequential effects: As we will see in a moment, an intermittently reinforced response usually takes longer to extinguish than a continuously (fixed-ratio one) reinforced response. But if an animal is trained first with intermittent reinforcement, then with continuous, and then reinforcement is withdrawn, the earlier experience is has an effect: the response extinguishes slowly, at a rate appropriate to intermittent reinforcement.

If, during initial training, reinforcement is predictable — delivered for every response, or at regular intervals, or only in the presence of a specific signal — then its omission will be easy to detect and extinction should be rapid. Conversely, if reinforcement occurs at unpredictable times, its omission will be hard to detect and extinction should be slow. For example, consider two groups of rats, one trained to run down an alley for food delivered on every trial (*continuous* reinforcement), the other trained to run to food delivered only on half the trials, randomly determined (*partial* reinforcement). How long will each group continue to run if food is omitted? It is hard to put oneself in a frame of mind where the actual result, the *partial* group runs longer, is surprising. Yet in the dawn days of learning theory, this unsurprising outcome was known as "Humphreys' paradox"; it is now known as the *partial-reinforcement extinction effect* (PREE). The paradox came from the early idea that reinforcement was invariably a "strengthener," which seemed to imply that the more frequently reinforced *continuous* group should have a stronger habit than the p*artial* group, so should run longer. As we have seen, animals adapt to reinforcement contingencies in much more subtle ways than the old strength model implies.

The effect of training procedure on resistance to extinction (persistence) is only one of the questions one might ask about extinction. Other questions concern the effects of training pattern on behavior in extinction and the effects of extinction on relearning of the same and related tasks; the first question gets at the processes involved in original learning ("what is learned," this topic has been much discussed in earlier chapters), the second can tell us something about the way the organism is changed by extinction.

Persistence is perhaps the most tractable problem; most work has been done on it, and the questions about it can be posed most clearly. Consider again the continuous-versus partial-reinforcement alley study. How much more effective is partial than continuous reinforcement in building persistence? The first question that comes up concerns the number of initial training trials: should we equate the two groups in terms of total number of trials — in which case the *continuous* group will get twice as many reinforcers; or in terms of number of reinforcers — in which case the *partial* group will get twice as many trials? The first choice is the usual one, but there is obviously no right answer to this question. A better one may be: What is the effect of number of reinforcements on resistance to extinction? If we can't control for something, then just take the bull by the horns and measure it directly. This was done in several classic studies,[8] with mixed results. The earliest studies seemed to show a monotonic effect: the more rewards received, the greater subsequent resistance to extinction. But later work sometimes suggested a nonmonotonic effect: resistance to extinction rises at first, but then declines in very well-trained animals. And of course, in these studies, number of *trials* is perfectly confounded with number of reinforcements; to attempt a deconfounding would just bring us back to the partial vs. continuous reinforcement experiment. Evidently there is no solution to the problem here.

An alternative tack is to look directly at the effect of the training procedure (number of trials, frequency of reinforcement) on the level of *behavior* — speed of running, rate of lever pressing, probability of a conditioned response. This is a familiar problem: in free-operant experiments, the relation between frequency of reinforcement and rate of response is the *response function*, discussed extensively in earlier chapters. Similar functions can be obtained for any response and reinforcement measures. If we restrict ourselves to *asymptotic* behavior, that is, behavior after sufficient training that it shows no systematic change, response functions are stable and well defined. For example, for variable-interval reinforcement, over most of the range the response function is negatively accelerated: as obtained rate of reinforcement increases, response rate increases, linearly at first, and then at a decelerating rate up to a maximum. For ratio schedules, over a similar range, the function has a negative slope, for spaced-responding (DRL) schedules it is linear, and so on (see Chapter 7).

We can get at the resistance-to-extinction problem by considering the effective *stimuli* controlling this behavior. There are two kinds of discriminative stimuli: the reinforcer itself (i.e., food, shock), and everything else. We have already seen considerable evidence for the discriminative function of reinforcers, and there is additional evidence. For example, if a food-reinforced activity is thoroughly extinguished, and then free food is given to the animal, responding usually resumes at once. If the temporal pattern of reinforcement is maintained in extinction, but on a response-independent basis (e.g., a variable-time schedule for animals trained on variable-interval), extinction is greatly retarded. This persistence is evidence both that the role of response contingency is minimal once the effective response has been acquired and that periodic food delivery plays a major role in sustaining the instrumental response. Even stronger proof of both these points is provided by a number of studies comparing persistence of a group of animals trained with an omission contingency (a food-avoidance schedule similar to shock postponement) in extinction with yoked animals that receive the same number and distribution of food deliveries, but independently of responding (Rescorla & Skucy, 1969; Uhl, 1974). The general result of these experiments is that the rate of extinction is the same in both groups; in other words, it is the absence of food as a stimulus (rather than absence of the response-contingency) that is responsible for most, perhaps all, the reduction of responding in extinction. I described earlier similar results from experiments on autoshaped pecking. The presence or absence of reinforcers, quite apart from any response contingency, is the major variable that determines persistence.

Suppose we assume, in line with the argument in Chapter 11, that the two sources of stimulus control, reinforcement and everything else, are additive. These assumptions can be summarized thus:

$$x = aR(x) + S_x, \tag{16.1}$$

where x is the rate of response X, $R(x)$ is the reinforcement rate for X, S_x is the (nonreinforcement) discriminative stimulus for X and a is a constant proportional to reinforcement magnitude representing the relative contribution of reinforcement and stimuli other than reinforcement to discriminative control of X: this formulation assumes that the contribution of reinforcement to the total stimulus control of x is directly proportional to reinforcement rate. x is related to $R(x)$ by the response function f: $x = f(R(x))$. Hence the reduction in x associated with omission of reinforcement ($R(x) = 0$) is given by

$$\Delta x = f(R(x)) - aR(x), \tag{16.2}$$

which is a reasonable first-approximation estimate for resistance to extinction: directly proportional to base rate of behavior, and inversely related to base rate of reinforcement -- I give a slightly more rigorous derivation of equation 16.2 in the Notes.[9]

These two assumptions are illustrated graphically in Figure 16.6; the four data points are from a rabbit nictitating-membrane classical-conditioning experiment by Gibbs, Latham and Gormezano (1978) but the general form of the response function is a familiar one. The straight line through the origin is just the function $x = a\,R(x)$, which is the hypothesized contribution of reinforcement to total response rate according to my descriptive model. The difference between these two functions (e.g., line segment AB for the 50%-reinforcement condition) represents the net response strength at the outset of conditioning, i.e., our estimate of persistence. Note the ranking of these distances for each reinforcement-percentage condition: 25>50>15>100. For the most part, resistance to extinction is predicted to be greater the less frequent reinforcement during training; there is a partial reversal for the 15% condition, reflecting the relatively low base response probability during training. This apart, it is clear that this approach does predict the partial-reinforcement effect.

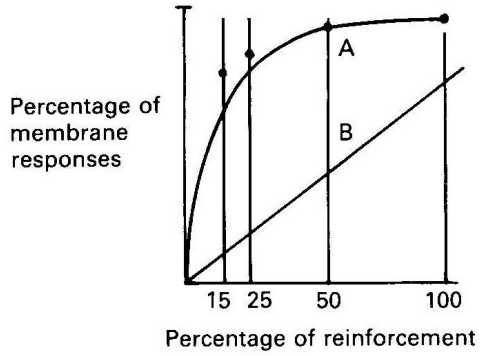

Figure 16.6. The negatively accelerated curve is the estimated response function relating percentage of conditioned responses to percentage of reinforcement from a rabbit-nictitating-membrane conditioning experiment by Gibbs, Latham, and Gormezano (1978) – the four group-data points are shown. The ray through the origin represents the hypothetical contribution of reinforcement to total response strength, according to the descriptive model discussed in the text.

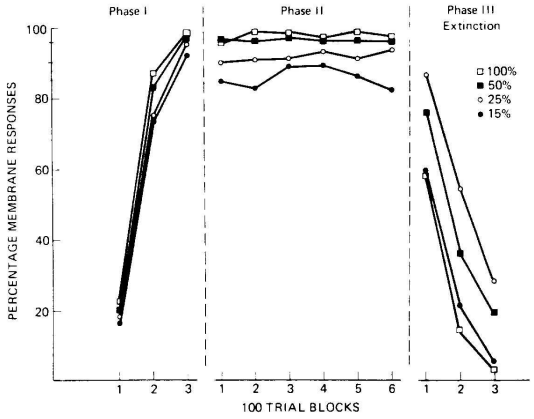

Figure 16.7. Three phases of a classical- conditioning experiment with the rabbit-nictitating-membrane response. Phase I (left): acquisition with 100% reinforcement. Phase II (center): maintenance with 100, 50, 25, or 15% reinforcement. Phase III (right): extinction. (After Gibbs, Latham, & Gormezano, 1978.)

Figure 16.7 shows the rest of the data from the Gibbs et al. experiment. The average maintenance levels in Phase II (maintenance) are shown across trials in the center panels, and as

overall averages in the previous figure. The left panel in Figure 16.7 shows the changes during acquisition of the membrane response, the right panel shows the changes during extinction. There is a PREE, in the sense that the 100% group extinguishes most rapidly, the 25% group least rapidly. The rank ordering in the right panel of Figure 16.7 is the same as the ordering just derived from Figure 16.6.

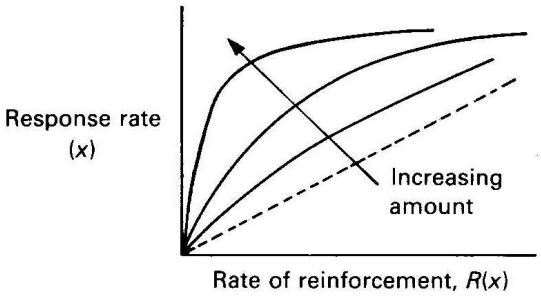

Figure 16.8. Effect of amount of reinforcement on a typical negatively accelerated response function.

This descriptive approach can handle a number of variables that have been shown to affect persistence, such as species, type of response, and amount of reinforcement. For example, for a given reward frequency, resistance to extinction is reduced by increasing reward size. The constant, a, in equation 16.2 represents the relative contribution to response strength of reinforcement factors, the slope of the straight line in Figure 16.6. Increasing reward size increases a, and thus term $aR(x)$, in direct proportion. The term $f(R(x))$ is also increased, but because f is a negatively accelerated function, this increase will always be less than proportional. Hence, the net effect will be to reduce Δx, our estimate of resistance to extinction, which is the usual empirical result. This result is intuitively plausible, in the sense that the change in extinction from some reward to none should be more easily discriminated if the original reward is large.

The model assumes that reward amount has two effects: to increase a, and to increase the rate of approach of the response function to its fixed maximum. The first assumption I make here for the first time; but the second assumption follows from optimality principles discussed earlier. The effect of reward amount on response rate is shown in Figure 16.8. The dashed straight line is the contribution to response strength of reinforcement rate, as before. The figure shows that at low amounts of reinforcement a PRE should be much more difficult to obtain than at high amounts. The reason is that the PRE depends upon the curvature of the response function, almost independently of the slope of the straight line representing the contribution of reinforcement rate or probability to response strength. Since reducing reinforcement amount reduces the curvature of the function (see note 10 and Chapter 12), the PRE should be correspondingly reduced. This effect of large rewards in promoting the PRE is well known.

It is likely that this descriptive approach can also account for many species and response-type differences. For example, response types differ in their sensitivity to reinforcement. Pecking and lever pressing are very sensitive and occur at high rates, even at low levels of food reinforcement. On most schedules, their response functions rise steeply at first and then flatten out; they show considerable curvature, and therefore the PRE is easy to obtain. Other responses, such as classically conditioned salivation, or treadle-pressing for food by pigeons, show a more proportional relation to reinforcement probability, that is, their response functions are closer to being linear. Hence, a PRE is difficult or impossible to demonstrate. Swimming by goldfish is perhaps more like treadle pressing than key pecking, hence it not surprising that a PRE has proven difficult to find, except when reward magnitude is large.

Under appropriate conditions, response functions of almost any form can be found. For example, consider the concurrent variable-interval situation discussed extensively in Chapter 8: The animal responds to two keys, each providing food on independent VI schedules. If the total reinforcement rate is held constant ($R(x) + R(y) = K$), then response rate on one key should be directly proportional to reinforcement rate on that key.[10] We would predict, therefore, that there should be no partial reinforcement effect here. If food is discontinued for responding on one alternative, responding should cease sooner the lower the reinforcement rate in training. Unfortu-

nately, precisely this experiment has not been done, but the prediction certainly accords with one's intuition that when another choice is available, a less-frequently-reinforced choice should be more readily given up.

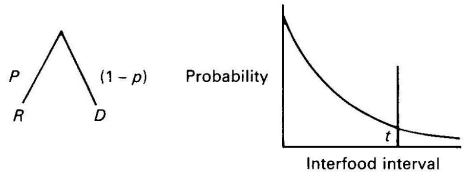

Figure 16.9. *Left*: Decision tree for extinction. *Right*: distribution of interfood times on a random-interval schedule.

This approach seems to apply equally well to positive and negative reinforcement. We have already seen that avoidance behavior is highly persistent in well-trained animals. Since shock occurs rarely under training conditions, it can make little contribution to response strength; hence its absence in extinction results in little reduction in the animal's tendency to respond, and behavior persists. The greatest persistence has been shown in experiments with variable-interval shock postponement. In this procedure, in the absence of responding, brief shocks occur at an average interval t_S; if a response occurs, the next shock occurs at an average time t_R where $t_R > t_S$. If $t_S = t_R$, there is no benefit to responding and the situation is analogous to the delivery of free food in extinction. Rats well trained to avoid take a very long time to cease responding under these conditions, attesting to the importance of shock as part of the stimulus complex sustaining avoidance responding.

Resistance-to-extinction experiments are expensive to carry out because a separate group of animals is required for each condition. A theory that depends upon parametric information about response functions is therefore difficult to test, and this one has not been adequately tested. Nevertheless, it summarizes much of a very large and confusing literature and is worth presenting on that account. I hope that more adequate tests will be forthcoming.

Extinction provides an animal with a detection task. Consequently, optimality theory is a natural way to look at resistance to extinction. Consider an animal on a variable-interval schedule. After much training, the feeder mechanism is turned off: When should he quit? The diagram on the left in Figure 16.9 shows a simple way of representing animal's problem. He has two hypotheses to consider: (a) That the VI schedule is still in effect, but he has encountered an extra-long interfood interval, or (b) that the VI schedule is no longer in effect. If (a) is true, then he can expect to get food with a frequency *R, where R* is 1/VI value. If (b) is true, then he can expect to get food at some residual rate, D, depending on his priors and past experience.

The mutually exclusive probabilities p, and 1-p associated with these two hypotheses depend upon the distribution of interfood intervals the animal has experienced in the past in the apparatus. Since we are assuming an experienced animal, these are given by the actual distribution of interfood intervals prescribed by the VI schedule. This distribution, for a random-interval schedule, is shown on the right. The area to the right of the vertical line at time t, labeled $p(>t|A)$, represents the proportion of intervals longer than t. If we time extinction from the last food delivery, then the animal must estimate p, the probability of hypothesis A, given a time t since the last food delivery. This can be done theoretically using Bayes rule, which gives $p(A|t)$ as a function of the animal's prior estimate of A (initially close to one), $p(>t|A)$, and $p(>t|B)$ (which equals one, since no food is delivered if B is true). The problem is tricky, because the analysis must take into account that $p(A|t)$ is constantly updated as t increases.[11]

Despite the formal complexities, it is easy to see intuitively that the animal's estimate of $p(A|t)$ must slowly decrease as t increases, because it becomes less and less likely that an interval this long without food could occur under the prior VI schedule. Thus, the animal should increasingly spend its time doing things other than making the previously reinforced response, leading to the typical negatively accelerated extinction curve. Analyses of this sort can be carried out for any instrumental conditioning arrangement. The virtue of this type of analysis, pursued rigorously, is that it can provide precise predictions not only about the form of the extinction curve, but also about the resistance to extinction to be expected from different types of schedules, that

is, different distributions of interfood intervals, response ratios or trials. Violations of these predictions point to constraints that can provide useful information about learning mechanisms.

SUMMARY

This chapter analyzed a number of familiar learning experimental situations — chained schedules, conditioned emotional response, avoidance and escape, set, response-produced electric shock and so-called "learned helplessness," extinction, and others from what might be termed a Darwinian perspective. Darwinian in two senses: first in the sense that operant learning is viewed as the outcome of an interplay between selective forces — reinforcement and punishment — which act mostly through temporal contiguity, and creative or variational forces that originate to-be-selected behavior. And Darwinian in the traditional sense that behavior in the organism's "selection environment" (or "environment of evolutionary adaptation" in current jargon) subserves Darwinian fitness: it is (usually) adaptive. The variational forces for learning include so-called "cognitive" processes that allow the animal to categorize situations in various ways. The scare quotes around "cognitive" are because the most important ways that animals categorize relate not to cognition but to emotion and affect, i.e., they are motivational, not informational in an abstract sense.

My theoretical aim has also been Darwinian. To suggest simple explanations that assume little but apply as broadly as possible. Thus, numerous phenomena related to resistance to extinction seem to follow from simple ideas of stimulus additivity and the form of the function relating response strength to reinforcement variables. Many of the properties of chained schedules follow from ideas of temporal control by reinforcement and non-reinforcement stimuli. Paradoxical effects of strong reinforcers and punishers follow from the expected effects of experience with such stimuli on the range of behavioral variation.

But these ideas provide little more than a framework at this stage. Much more theoretical and experimental work will be needed to understand the dynamic processes that produce the wonderfully regular experimental data now available on the operant behavior of animals at a level of detail that begins to allow some connection with the underlying neurophysiology.

The ethologist Tinbergen (1963) described four questions that can sensibly be asked of behavior: (a) Its selective value or function; (b) its causation (controlling stimuli and motivational factors); (c) its development (ontogeny); and (d) its evolutionary history (phylogeny). These questions can be asked of any species-typical behavior, whether learned or innate. I have discussed behavior from all four points of view in this book, although the emphasis is on the first three.

The study of learning mechanisms combines two of Tinbergen's factors: The study of causation, of controlling stimuli, is the study of "what is learned," in psychological terminology. But the study of the *process* of learning is essentially developmental. Operant learning is not (as was once widely believed) a gradual accretion of "strength" by a stimulus-response system akin to the reflex, but rather the building up of a program for action, with inputs both from the present environment and a representation of the past. The study of the learning process is therefore more a question of identifying necessary stages through which this process of representation-building and program assembly must pass, than the tracing out of some smooth curve. As in development, environmental feedback is involved at each stage. For example, the development of operant pecking involves first learning of predictive stimulus-reinforcer relations. This is usually termed *classical conditioning*, but it has little in common with reflex strengthening: The animal is learning what leads to what, not building a mental muscle. Under uncontrolled conditions, this learning is inextricably mixed up with the occurrence of exploratory responding and the stimulus changes it produces. The separate role of the process of "representation building" only became clear when pecking was studied in open-loop situations where it was permitted no effect. Then the two-stage process was revealed: first the bird learns what predicts food, only then does it be-

gin to peck. We know rather little about the further stages, except that selection mechanisms, involving both the temporal (predictive) relations between activities and reinforcers and priors derived from heredity and past experience, act to winnow out ineffective variants from the pool provided by classical conditioning.

Future work proceeds on two fronts — which roughly defines the division between classical and operant conditioning: to understand more about the properties of animals' representations, how they are formed, and what they are: What do animals know, and how do they learn it? And to understand more about the selection of responses: What determines the pool of variants, and how are some activities selected over others?

NOTES

1. *Concurrent-chained schedules*. There has been considerable interest in chained schedules over the years because they seem to offer a way to measure reinforcing value independently of reinforcement schedule. A typical concurrent chained schedules comprises two independent, two-link chained schedules in which the first links are usually the same (typically VI 60-s schedules) and the second links are different. Reinforcement for pecking a first-link choice is the appearance of the second-link choice on that key; the other choice is disabled until the animal collects one or a few second-link food reinforcers, after which both first links reappear. Typically, first-link reinforcements are programmed by a single interval timer, with random, equal assignment to each side. The second-link food reinforcement is separately programmed for each key, by interval timers that run only while that component is in effect.

The attraction of this procedure is the notion that the proportion of responses on each of the first-link choices can provide a measure of the conditioned-reinforcing effectiveness of the second links that is independent the rates and patterns of responding maintained by the second-link schedules. That promise has not really been fulfilled, because the strength concept of reinforcement has not lived up to expectations, and because conditioned reinforcement, in particular, has turned out to be a more complex idea than its name suggests.

The concurrent-chains procedure is complicated. The second-link contingencies favor exclusive choice: Since each timer runs only so long as the animal is in that component, there is no point entering the component with the longer average time to food. This expectation is borne out by experiments in which separate VIs (rather than a single timer) were used for the first link: if the first links are relatively short, pigeons tend to fixate on the key with the better second link. But the more typical, single-assignment first-link VI favors nonexclusive choice, since it ensures that fixation on one key will soon lead to extinction (see Chapter 8, and Staddon, Hinson & Kram, 1981). It is not easy to anticipate the resultant of these two conflicting contingencies. The results of concurrent-chained-schedule experiments are accordingly less reliable than the results of the simple concurrent studies discussed earlier.

In an early experiment, Herrnstein (1964) found that the relative frequency of pecking on the two first-link VIs matched the relative rates of reinforcement in the second link VIs; for example, the birds would peck 2:1 on the left key in the first link if the two second link VIs were VI 30 sec and VI 60 sec. This seemed an encouraging confirmation of the generality of the *matching law* (see Chapters 8 and 11) and supported the view that the procedure could provide a way of measuring the effectiveness of conditioned reinforcement.

Later work has shown that Herrnstein's result is not general. More commonly, if the first link is not too long relative to the second links, pigeons tend to *overmatch*, that is disproportionately favor the better second-link VI: If the second-link VI reinforcement rates are in the ratio 2:1, choice proportions might be in the ratio 3:1 (e.g., MacEwen, 1972). In addition, first-link choice depends upon the absolute values of both first and second links. For example, Fantino

(1969) showed that preference for the richer of the two terminal VIs diminishes as the length of the first-link VIs increases. This makes adaptive sense: If the first-link VIs are much longer than the second-link VIs, then the animal's objective should be to get into the *either* second link, since both are so much closer to food than either first link. Hence, the longer the first links relative to the second links, the closer to indifference should the animal become.

An optimality analysis of concurrent-chain schedules with independent (i.e., dual assignment, not single-assignment) VIs in the first link is in fact quite straightforward. There are only two feasible strategies: either respond exclusively to one side, or respond to both (i.e., alternate between the first links so as to pick up either second link when it becomes available). The analysis, an extension of the analysis of delay situations in Chapter 8, is as follows: The expected food rate, given exclusive choice and assuming that each interval schedule runs only when its stimulus is on is

$$1/(T_1 + t_1), \qquad \text{(N16.1)}$$

where T_1 is the interval value in the first link, and t_1 the interval value in the second link. If the animal approximately alternates in the first link so that either second link is entered as soon as it becomes available, then the expected food rate is the sum of the average time in the first link, plus the expected time to food in the second links, weighted by the proportion of time the animal enters each of the second links. The average time in the first link is then the reciprocal of the sum of the rates, that is, $T_1T_2/(T_1+T_2)$. The average time in the second link is just $t_1p + t_2(1-p)$, where $p = T_2/(T_1+T_2)$. If alternative 1 is the majority choice, then the switching condition for responding exclusively to 1, versus responding to both, is the sum of these two equations, versus Equation N16.1, which boils down to

$$1/(T_1 + t_1) \geq (T_1 + T_2)/(T_1T_2 + T_2t_1 + T_1t_2), \qquad t_1 < t_2$$

which reduces to

$$t_2 \geq T_1 + t_1. \qquad \text{(N16.2)}$$

Equation N16.2 has the surprising feature that the decision whether to respond to one or both alternatives depends on only one of the first-link interval values: the value associated with the shorter of the second links. The reason is that after a certain amount of time without choosing alternative 2, the expected time to enter its second link will be close to zero; hence it always pays to choose 2 at some time, so long as $t2$, the expected time to food in its second link, is less than the expected time to food on side 1, which is the sum of the two expected delays on that side.

Equation N16.2 can be represented in the usual way as a switching line in t_1/t_2 space, as shown in Figure 16.10. The results of two kinds of experiment can be predicted at once from the figure. First, increasing T, the length of the initial links, moves the switching line up the t_2 axis, and thereby increases the range of t_1 and t_2 values for which the animal should choose both alternatives rather than fixating on one. As this would lead one to expect, pigeons are increasingly indifferent between the two alternatives as T increases. The second prediction is indicated by the dashed line in the figure, which represents the case where $t_1<t_2$ and t_1/t_2 is held constant while the absolute values of t_1 and t_2 are both increased. Clearly as the absolute

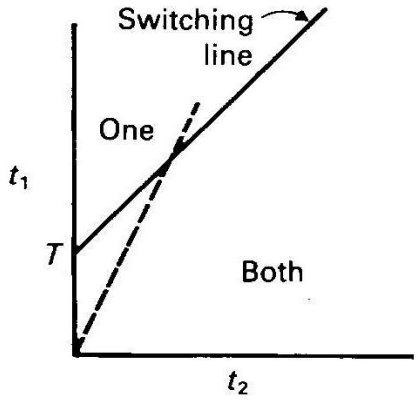

Figure 16.10. Switching-line analysis of concurrent-chained schedule. Solid line is Equation N14.2 in the text, with $T_1 = T_2 = T$ (i.e., equal first links). The dashed line shows the effect of increasing the absolute values of the second links (i.e., t_1 and t_2) while maintaining their ratio constant.

values of t_1 and t_2 are both increased, preference should shift towards the shorter second link, and this has also been found (cf. MacEwen, 1972; review in Fantino, 1977).

Equation N16.2 can also predict what should happen when the first-link VIs are varied. For example, suppose both first-link VIs are reduced in value while their ratio remains the same: what will be the effect on preference? If T_1 becomes negligible, then Equation N16.2 reduces to $t_2 \geq t_1$, that is, exclusive choice of the shorter second-link VI. (It may surprise that the optimal solution does *not* converge on an approximation to matching as the first link VIs are reduced in value, but of course the reason is that the second link VIs do not run concurrently — hence there is no point entering the longer second-link VI.)

The simple switching-line analysis cannot account for variations in the degree of preference within the "both" region, although it is obviously possible in principle to extend it in several ways, none of them inviting. A few descriptive models of the matching-law variety have been proposed. The simplest is owing to Fantino (e.g., 1981), as follows:

$$x_1 / x_2 = (D - t_1)/(D - t_2), \quad t_1, t_2 < D \qquad \text{(N16.3)}$$

where D is the overall average time to food at the beginning of a cycle (onset of the first link), and x_1 and x_2 are rates of response in the first link. If either t_1 or $t_2 > D$, then exclusive choice is predicted. This model can both explain the effects of increasing T and of increasing the absolute values of t_1 and t_2 while leaving their ratio constant.

A weakness of Equation N16.3 is that it predicts indifference whenever $t_1 = t_2$, irrespective of the value of the first links. Squires and Fantino (1971) therefore proposed a modified version where the delay-reduction ratio (the right-hand side of Equation N16.3) is weighted by the overall food rate on each side:

$$x_1 / x_2 = (D - t_1)R_1 /(D - t_2)R_2, t_1, \quad t_2 < D, \qquad \text{(N16.4)}$$

which ensures that the equation converges on matching as the second-link VIs approach zero. Killeen (1982) has extended his arousal model to a wide range of choice procedures, including concurrent chains. Fantino's model and any optimality analysis, is parameter free, but says nothing about the means animals use to adapt to these procedures. Killeen's model has two parameters, but promises to relate choice to processes of arousal. Killeen and Fantino reconciled the two models in 1990.

A parenthetical note on the virtues of optimality analysis: It took several years after the first experiment with the concurrent-chains procedure before someone noticed that the value of the first-link VI schedules should make a difference to preference. It took a few more years before they noticed that the values of the first-link VIs need not be the same, and that this also might make a difference to choice. There is still no clear distinction drawn between the results to be expected from single-assignment (sometimes called *interdependent*) first-link VIs and dual-assignment (independent) VIs. All these factors come to immediate attention as soon as we ask: What is the optimal strategy? Hence, a major virtue of optimality analysis, quite apart from its theoretical merits, is that it draws attention to what are very likely to be the critical experimental variables.

2. Reviews of experimental work on positive conditioned suppression appear in Lolordo, McMillan and Riley, 1974, Schwartz and Gamzu, 1976, and Staddon, 1972. See also Buzsáki (1982) and Buzsáki, Grastyán, Winiczai, and Mód (1979).

3. This experiment and the lever-press studies discussed here are reviewed in Hineline (1977).

4. For example, in delay studies with positive reinforcement, it is always necessary to decide what to do about responses after the effective response (i.e., the response that starts the delay timer that produces reinforcement): should they be allowed to occur, in which case the reinforcer

may follow some responses after less than the prescribed delay? Or should later responses cancel or put off reinforcement, in which case the delay contingency may have a substantial effect on reinforcement *rate*? The compromise of retracting the lever or preventing later responses in some other way introduces an additional stimulus, with its attendant problems of behavioral reallocation. These same problems would confront any attempt to measure a delay gradient for negative reinforcement.

5. It is worth noting that the present treatment of reinforcement in terms of selection from among a stimulus and response candidate set avoids the problems posed for traditional reinforcement theory by avoidance procedures. The creation of the candidate set via surprise, arousal and inference mechanisms (termed *mechanisms of behavioral variation* in an earlier account, see Staddon & Simmelhag, 1971) has priority over selection. Consequently, the absence in shock postponement schedules of contiguity between response and a stimulus signaling shock reduction just emphasizes the role of inducing mechanisms in the behavior. And as we have seen, these mechanisms are indeed of primary importance in avoidance: the type of response required is critical, as is the animal's past history and the intensity and frequency of shock. Contiguity is probably involved in the sense that the behavior least contiguous with shock tends to be favored, but its effect is indirect and subordinate to inducing factors.

6. *Cognition and learned helplessness*. There has been some controversy over whether the learned-helplessness effect represents a *cognitive* or *associative* deficit — or just an effect of conditioned passivity. There is perhaps less to this argument than meets the eye. Few now believe that all the effects of past experience on an animal are entirely manifest in its overt behavior. Hence all learning by mammals and birds is "cognitive," if by that term we mean nonmotivational internal changes more complex than links between simple stimuli and responses. The argument therefore boils down to a difference of degree: *How much* of the effect is explainable solely by motivational or activity-level changes? The answer is far from clear. Some authors are not convinced that any of the animal results require more than this; all agree that many of them can be explained in this way. A set of carefully controlled transfer experiments with rats seems to provide the best evidence for some kind of generalized "associative deficit" (Alloy & Seligman, 1979).

The vigor of the controversy perhaps derives not so much from unthinking behaviorists' disbelief in cognition, as from skepticism about the particular kind of rationalistic cognitive account most frequently offered. For example, dogs are said to have "acquired the expectation that shock termination would be independent of their responses." This profundity is termed "helplessness theory." The emptiness of the account is obvious once we replace the word "dog" with "computer." Given this account of a computer program, would we feel confident of understanding how it works? Such a statement simply restates the problem; the questions of how the animals learn, what they learn, and what general principles apply to the learning, remain. We saw in the last chapter that the term *expectation*, convenient though it is (and I have made use of it several times in this book), implies some form of representation; without some attempt to specify that representation, the word has no scientific meaning. Regrettably, almost none of the learned-helplessness research has begun to approach this question. Instead ever-more ingenious experiments obsessively flog once again the dead horse of a stimulus-response account now believed by almost no one. The answer to stimulus-response theory is not yet another control group, but some positive evidence on the form of the animal's representation. Data and theory on this point are sadly lacking.

7. For reviews of the fascinating recent work on human decision making see Einhorn and Hogarth (1978), Kahneman and Tversky (1979) and the book edited by Wallsten (1980); see also Gigerenzer and Selten (2002).

8. See the books by Mackintosh (1974), Osgood (1953) or Hulse, Deese and Egeth (1975) for good reviews of these old studies, and of the whole topic of extinction and partial reinforcement.

9. The status of reinforcement rate as a discriminative stimulus is a little different from the status of the usual synchronous stimulus, present all the time. Part of the problem is that the notion of discriminative stimulus is not itself well understood. I am inclined to think of a discriminative stimulus as akin to a memory-retrieval "address" that reinstates a program assembled during previous learning to deal with the situation identified by the stimulus. This is speculative however, and the direct action of most discriminative stimuli makes it easier to think of them almost as elicitors of the behavior under their control — although Skinnerian terminology avoids this by means of the elliptical usage that the discriminative stimulus "sets the occasion for" the operant response.

Since reinforcement occurs only episodically in most operant procedures, reinforcement rate cannot usually act directly to produce responding. Instead it may be that the synchronous stimuli act in some way to reinstate the action program, which also contains some representation for the expected frequency and distribution of reinforcement. Resistance to extinction is then inversely related to the *discrepancy* between the actuality which, in extinction, contains no reinforcement, but does contain the nonreinforcement discriminative stimuli, and the representation, which contains both: the smaller the discrepancy, the greater the resistance to extinction. This situation can be symbolized as follows:

$$S* = aR(x) + S_x, \tag{N16.5}$$

and

$$S = S_x, \tag{N16.6}$$

where $S*$ is the representation built up during training, S the situation in extinction, S_x the contribution to $S*$ of nonreinforcement factors and $aR(x)$ the contribution of reinforcement factors; a is a constant, so equation N16.3 assumes that the contribution of reinforcement factors to $S*$ is proportional to the rate of reinforcement.

Representation $S*$ then generates a rate of responding, x, which follows the appropriate response function. When reinforcement is omitted in extinction, $R(x) = 0$ and S and $S*$ differ by the term $aR(x)$. Resistance to extinction presumably depends both on this discrepancy — the larger the discrepancy, the easier it is to detect the change, hence the faster the extinction — and on the base rate of responding — the higher the rate, the longer it should take for responding to disappear. Under training conditions, $S*$ is associated with a rate of responding x, which is related to $R(x)$ by the response function f: $S* \rightarrow x = f(R(x))$. A simple relation that incorporates the effect of both discrepancy ($S*\text{-}S = aR(x)$) and base rate ($x = f(R(x))$) on persistence (P) is therefore,

$$P = f(R(x)) - aR(x), \tag{N16.7}$$

which is equation 16.2 in the text.

10. See Herrnstein (1961). This result follows from the matching relations discussed in Chapters 11 and 12: $x = kR(x)/(R(x) + R(y)) = kR(x)/K$ when total reinforcement is constant. The effect of reinforcement magnitude on the curvature of the variable-interval response function can be derived as follows: The descriptive equation for the VI response function is $x = kR(x)/(R(x) + R(z))$, where $R(z)$ is the reinforcement for "other" behavior. The term $R(z)$ also gives the value of $R(x)$ for which x is half the maximum value; the smaller the value of $R(x)$ at which this occurs, the greater the curvature of the function over a given range. Since the func-

tion deals only with relative measures, increasing the magnitude of reinforcement for x is equivalent to reducing the value of the reinforcement for z, hence to reducing term $R(z)$*;* a reduction in $R(z)$ corresponds to an increase in the curvature of the response function. Hence, increasing reinforcement magnitude increases the predicted curvature of the response function.

11. McNamara and Houston (1980; see also 1999) have tackled this problem in a difficult theoretical paper.