8
CHOICE AND DECISION RULES

The study of choice is the study of the factors that make animals do one thing rather than another. In this broad sense all psychology is the study of choice. There is another, more conventional, meaning: *conscious deliberation,* as when we mull over alternatives. Shall we go out and get a pizza, or cook at home? Is it to be medical school, grind, and the big bucks, or genteel poverty and the spiritual satisfactions of renaissance history? No doubt conscious deliberation does occur; but it is inaccessible in animals, and its causal status even in people is by no means clear. Often the reason follows the choice ("rationalization") rather than the reverse, for example. There are many striking examples of the disconnect between introspective self-justification and the actual causes of behavior from the literature of brain damage (see, for example, Pinker, 2002, p. 43 et seq.). So I'm concerned here just with the first meaning: An animal "chooses" response A only in the sense that it does A, rather than B, C, or some other thing.

The previous chapter was an account of choice in this sense. Two activities were linked by a contingency, and we looked at how their frequencies, and the frequencies of other activities, changed in consequence. There are other, more explicit, procedures for studying choice, however. Instead of rewarding one response and seeing how its frequency changes relative to others, we can reward two (or more), and see how the animal allocates his effort between them. The term *choice* is generally reserved for situations of this special sort.

Choice experiments are done for the same reason as experiments in which only a single response is reinforced: to find the rules by which animals adapt to reward and punishment. Choice experiments simplify things by pitting two or more similar responses against one another, and by rewarding each with the same thing (although not according to the same schedule or in the same amount). When an unrestrained animal shifts from one activity to another, it is choosing between different things: Running is different in kind as well as amount from eating or grooming. These activities are imperfect substitutes for one another. In most conventional choice experiments, on the other hand, the outcomes (usually food) are qualitatively the same. What differs between alternatives is the *means by which* food is obtained.

For example, one of the oldest choice procedures is so-called *probability learning.* An animal such as a rat, a pigeon, or even a goldfish, is confronted with two alternatives, which might be two arms of a T-maze, two keys, or two platforms to which it can jump (goldfish are poor at the latter task, but do quite well at the other two). In the simplest version of this procedure, food is assigned to one or other alternative on every trial with unequal, but complementary, probabilities. Thus, food may be available for choice A on 75% of trials (A is termed the *majority* choice) and choice B on 25% of trials. The animal must choose not between eating and some other activity, but between eating with this or that *probability.*

This is an example of a *discrete-trials* procedure: The animal is not free to respond at any time, but must confine his choices to trials that are specified by the experimenter. Here is another example: A hungry pigeon is confronted with two pecking keys; the keys are only illuminated and effective during a trial. A single peck on a key suffices to turn it off and, if the response is "correct," operate the food hopper. After the response, with its positive or negative consequence, the lights go out or the opportunity to respond is withheld in some other way; this is termed the *intertrial interval* (ITI). After the ITI, the lights come on again, and the next trial begins.

Omitting the ITI, and allowing the animal to respond at any time, converts this to a *free-operant* procedure. In this case, the two alternatives generally differ in the reinforcement schedule associated with them. For example, a popular procedure is to look at the rates of key pecking when one key provides food according to one value of variable-interval (VI) schedule, and the

other key according to a different VI; this is termed a *concurrent VI-VI* schedule.

There are three ways to look at choice situations: To see whether the changes in performance over time fit learning models; to look at the steady state, in molecular or molar terms; and normatively, to see if animals do what they should to maximize reward or behave optimally in some other way. I briefly describe each of these approaches and then spend the most time on the last two: steady-state molar and molecular performance, and optimality analysis.

Suppose you have a hypothesis about the effect of each eating episode (reinforcement) on the tendency to repeat the response that led to it, a quantitative version of Thorndike's law of effect (see Chapter 5). With an assumption about how the animal *samples* the alternatives before it knows anything about them, you can make predictions about changes in the relative frequency of A and B choices over trials. This is the most profound, and most difficult, kind of analysis, because what is sought is a model of the process by which choices are made and change in frequency with experience — a model of the *acquisition* of behavior. Despite its difficulties, until quite recently this was the dominant theoretical approach to choice.

A second tack is to set aside the problem of acquisition and just look at the steady-state pattern of behavior: What *decision rule* has the animal arrived at? In the simple probability-learning experiment just described, for example, most animals eventually fixate on the majority alternative, choosing A on every trial. Here the decision rule is both obvious and trivial, but in more complex situations, where choice is nonexclusive, the rule may be quite complicated.

Historically, there has been controversy about describing the behavior of animals by decision rules. Objections are of two types: to the implication of conscious deliberation, and to neglect of the learning process that this approach implies.

Some people object to decision rules because they seem to imply conscious deliberation. This objection is heard less frequently as rule-following machines have become commonplace — computer scientists can speak of machines that make decisions without being seriously accused of mentalism or imputing consciousness to silicon chips. An animal may follow a decision rule as a baseball follows Newton's laws of motion, with little reflection.

The second objection is concerned with the process of learning and the properties of steady-state performance. A computer can be programmed to follow a rule, but an organism in a choice experiment must learn to apply and perhaps even develop the rule through learning. Some have felt that looking at the rule without worrying about how it comes about evades the problem of learning. But if animals show choice patterns that follow rules, realistic learning models must show how these patterns develop. Unfortunately, most learning models so far proposed are stochastic (probabilistic) and are not concerned with moment-by-moment behavior. For example, many assume that on each trial the animal chooses alternative A with probability p and B with probability $1 - p$. Each reinforced choice is then assumed to increment the probability of responding to that alternative: $p(n+1) = p(n) + \Delta$, where $p(n)$ is the probability of an A choice on trial n and Δ is the increment produced by reinforcement for that choice. Since probabilities must be less than or equal to one ($0 \le p \le 1$), Δ is usually defined in a way that limits the maximum value of p to unity, e.g., $\Delta = p(n+1) - p(n) = w(1 - p(n))$, $O < w < 1$, where w is a *learning rate* parameter. Correspondingly, on unreinforced trails, the increment is $p(n) = -vp(n)$, $0 < v < 1$, where v is the *extinction rate* parameter.

There is no room for a decision rule in such a stochastic model: The problem of moment-by-moment performance is solved simply by ignoring it. As we will see in Chapter 13, models of this type have been enormously fruitful through their predictions of molar steady-state performance. They are less useful as models of learning.

Lacking adequate learning models, we have little choice but to begin by looking at the end point, the properties of steady-state performance.

A decision rule is a *molecular* notion, it defines what the animal is doing moment by

moment. A less ambitious approach to steady-state performance is to look for *molar* rules that describe choice. We may be uncertain about what the animal is doing from moment to moment, but perhaps his average behavior — the proportion of majority choices, for example — follows some simple descriptive rule, an *empirical law* of choice. *Probability matching,* which states that the steady-state probability of the two choice responses in certain discrete-trial probability learning situations matches the reward probabilities, is one such principle. The *matching law,* which states that in concurrent VI-VI schedules the proportion of majority choices equals the proportion of majority reinforcements, is another. Probability matching, once thought universal, is in fact found rarely; matching does occur, under restricted conditions I discuss in a moment.[1, 2]

OPTIMAL CHOICE

The third way to look at choice is in terms of the optimal behavior implied by each procedure: we ask, not What does the animal do? but What should he do? For example, if the objective is to maximize average reinforcement probability, then in the simple probability-learning procedure after an initial sampling period there is no point in responding at all to the minority alternative. Hence, the typical steady-state pattern, fixation on the majority choice, is also optimal. Obviously, if most choice behavior is optimal in some straightforward sense, much mental effort can be saved: Instead of remembering the detailed results of innumerable experiments, we need record only the general rule they all follow. The study of optimal choice and the search for decision rules (or molar empirical choice laws) are therefore complementary: Optimality principles can lead us to decision rules, and make sense of decision rules already discovered. In the remainder of this chapter, I discuss choice from the point of view of decision rules and optimality.

Probability learning[3]

In simple probability-learning experiments, as we have seen, most animal species eventually follow the optimal strategy — exclusive choice of the more-probable-payoff (majority) alternative. This result implies a very simple general principle: At any point animals choose the best option available. Many find the simplicity of this rule appealing, but, alas, many others do not. In consequence, most attention has been devoted to procedures that yield nonexclusive choice, where even if this principle still holds, its operation is unobvious. These situations provide a graded rather than all-or-none dependent measure — choice proportion rather than an individual choice. They ensure that animals do not respond exclusively to one alternative or the other, but divide their responses between the two.

In what follows I attempt two things: explain what it is about these procedures that leads to nonexclusive choice, and show how the simple rule “at any point, choose the best option” — the hill-climbing rule discussed earlier in Chapter 2 — underlies many of them.

As we have already seen, when animals are allowed to choose, trial-by-trial, between two alternatives rewarded with different probabilities, most end up choosing the majority alternative on every trial. There are two main ways to modify this situation so as to produce nonexclusive choice: *hold* procedures, and *single-assignment* of reward. In *hold* procedures, a reward once made available for a response remains until the response occurs; in *single-assignment* procedures, a reward is made available for one response or the other and no other assignment is made until the assigned reward is collected. In the simple probability-learning situation, where reward is always available for one of the two choices, these amount to the same thing. It turns out that if food once assigned to an alternative remains available until the correct choice is made (and no new assignment is made until that happens), animals will persist in sometimes choosing the minority alternative, that is, nonexclusive choice.

The way in which an animal chooses from trial to trial depends upon the species and the individual. In single-assignment procedures, monkeys and sometimes rats will eventually develop a *lose-shift* pattern: After receiving food for a correct choice, the next choice is always of

the majority alternative. If this is correct, the cycle begins again; if not, the next choice is the minority alternative. Since food is always assigned to one of the alternatives on every trial, this minority choice is always reinforced, and the cycle begins again.

This pattern is clearly the most effective one open to the animals: After a rewarded trial, the probability of food is highest on the majority choice, which is the one chosen. But if the choice is unrewarded, food is certainly available for a minority response.

Probability-learning-with-hold illustrates two concepts that will be useful in future discussion: An *initializing event* (IE) is any signal available to the animal that tells him the state of the programming apparatus. The optimal sequence is perfectly defined from that point. The IE in probability-learning-with-hold is food delivery, since the reward probabilities are always p and $1 - p$ on the next trial. The *optimal sequence* is the sequence of choices following an IE that maximizes either momentary or overall payoff (here both). Here the optimal sequence is obviously AB, where A is the majority choice: Food delivery resets the sequence, so that response A always occurs after food; but if no food follows the A choice, the next choice is B, which is always rewarded, and the sequence begins again.

There are two kinds of optimal sequence, *local* and *global.* Local optimizing is just choosing on each trial the alternative with the highest probability of payoff. Other names for local optimizing are hill climbing, gradient descent and *momentary maximizing*. As we saw in Chapter 2, hill climbing doesn't always find the globally best behavior, but in most commonly studied choice situations, it does very well: The difference between the local and global maxima is usually trivial. In probability-learning-with-hold, the local and global optimizing sequences are the same. In most situations, picking the best option on each trial will come close to maximizing overall payoff rate.

While many animals learn to behave optimally in probability-learning-withhold, others do not. Goldfish, for example, do not develop the precise lose-shift pattern, although they do continue to respond to both alternatives, rather than fixating on one. Why not?

Learning an optimal choice sequence requires that an animal be guided by critical events, both external (IEs), such as food delivery, and associated with its own behavior, such as a particular choice. The decision rule for probability-learning-with-hold is "after food choose A, after an unsuccessful choice [which will always be A if the animal follows the sequence], choose B." Thus, the animal must be capable on trial n of being guided by what happened on trial $n - 1$. The ability to behave optimally therefore depends on the properties of *memory.* It is very likely that the inability of some animals to follow the lose-shift strategy in probability-learning-with-hold reflects memory limitations: The animals are confused about what happened on the previous trial, hence cannot use this information to guide their behavior on the current trial. Memory is a *constraint* (in the sense of the last chapter) that limits their ability to behave optimally.

We are not certain how memory limitations prevent goldfish from following the lose-shift pattern while permitting them to respond to both alternatives in the probability-learning-with-hold procedure. But the following scheme is a reasonable guess: Suppose that the animal is really a hill climber — it always goes to the place where food is most likely. To follow this strategy requires the animal to remember when it last got food: Was it on the previous trial, or an earlier trial? If it ever gets confused about this, it will respond incorrectly. For example, if it *didn't* get fed on the previous trial, but remembers it did, it will choose the majority rather than the minority. Conversely, if it did get fed, but remembers that it didn't it will choose the minority rather than the majority. But if these memory errors are at all frequent, the animal has no opportunity even to learn the optimal strategy in the first place. Consequently, its choice pattern will be a moment-by-moment mixture of *extinction* and *sampling:* It chooses some alternative where it has been rewarded, and then, if reward happens not to be available on that trial, persists for a while until its memory of past rewards for that choice grows dim, whereupon it chooses the other. With lapse of time, the first choice grows more attractive (so-called *spontaneous recov-*

ery), and reward may not be available for the second choice — both factors will favor a return to the first choice at some time. Added to these systematic effects is the unsystematic effect of sampling (i.e., unpredictable variation) built in to ensure that the animal does not get trapped (cf. the discussion of behavioral variability in Chapter 3).

If we look just at choices after rewarded trials, animals that follow the optimal sequence look different from animals that do not. Animals that behave optimally always respond to the majority choice after a rewarded trial; whereas animals that do not will divide their choices between the two alternatives. Often they divide them approximately in the same proportion as the rewards for the two choices (probability matching), but more usually, the majority choice is disproportionately favored. Goldfish tend to match, monkeys to respond exclusively to the majority. Monkeys were thus said to maximize, goldfish to match, but this is misleading. The argument I have just made suggests that both species are maximizing, but because of their different memory capabilities, they perform differently. There is no such thing as maximizing in an absolute sense; the term makes sense only in relation to a set of *constraints.* If the constraints are different so, usually, will be the behavior.

Whatever the details, it is clear that the hold procedure will act to maintain nonexclusive choice, even in an animal lacking the capacity to learn the optimal sequence. Consequently, the resulting choice sequence might well lack any simple pattern. Other limitations come from response tendencies that reflect particular apparatus features. For example, rats tend to pick a choice other than the one that was just reinforced (after all, in nature, once you have eaten the food, it is gone and usually will not soon reappear). Inability to overcome this tendency may account for failures to observe optimal sequences in some animals.

The general point is that either momentary maximizing under discriminative stimulus control, or cruder mechanisms involving extinction of unrewarded responses, will tend to produce nonexclusive choice in *hold* procedures. In more complex choice procedures where food is assigned to one or other alternative, but then only made available for a choice with some probability, the optimal pattern hovers around *matching* of choice proportions and payoff proportions. I return to matching in a moment.[4]

Delayed outcomes: "self-control"

In simple probability-learning procedures, reward immediately follows the correct choice. All that varies from condition to condition is the probability of reward for each choice. A modification that sheds considerable light on the nature of choice is to make reward available for either choice on each trial (i.e., abolish the probabilistic aspect), but vary reward amount and add a variable amount of delay to each choice. Such a procedure is illustrated in Figure 8.1. It has three main features; an intertrial interval (ITI), two reward amounts, R_1 and R_2, and two associated delays of reward, t_1 and t_2.

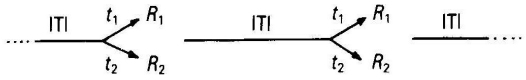

Figure 8.1. Procedure for studying the effects of delay and reward magnitude on choice. After an intertrial interval (ITI), the animal can choose between a small reward, R_1, with delay t_1, or a large reward, R_2, with delay, t_2.

Procedures of this sort are sometimes studied because of their resemblance to human situations involving delayed gratification ("self-control"). If R_1 is larger than R_2, and t_1 is not proportionately larger than t_2, then in some sense it "pays" to choose the alternative with the longer delay, because the larger outcome more than compensates for the longer delay. For example, suppose R_1 is 6-sec access to grain and R_2 is 2-sec access, but t is 4 sec whereas t_2 is 2 sec. Obviously, the expected rate of food delivery at a choice point is 1.5 sec/sec (6/4) for the large reward but only 1 sec/sec (2/2) for the small. In one sense, therefore, the "rational" rat should delay gratification, forego the immediate, small reward, and go for the larger, delayed one — thus sat-

isfying both enlightened self-interest and the Puritan ethic.

I'm not sure whether this arrangement has anything special to tell us about human delay of gratification, but animals exposed to it do behave more or less as this analysis suggests. For example, experiments with several variants of the procedure[5] have shown that if constant amounts of time are added to both t and t_2, preference can be made to shift: When t and t_2 are both short, the short-delay, small reward is preferred, but when t and t_2 are both long, the longer-delay, larger reward is preferred. This follows because the expected food rate depends upon the proportions, R_1/t_1 versus R_2t_2 , and these change when constants are added to the times.

The formal analysis is as follows: If future rewards are not discounted (i.e., valued less the more delayed they are), the animal should be indifferent between the two alternatives when

$$R_1/t_1 = R_2/t_2, \quad R_1>R_2, \tag{8.1}$$

that is, when the expected food rates are the same for each choice. Equation 8.1 can be rewritten as $t_1 = t_2(R_1/R_2)$, which defines a *switching line,* relating t_1 and t_2, shown in Figure 8.2. The slope of the line is the ratio of reward magnitudes, R_1/R_2, and the line divides the t_1-t_2 space into two regions: Above and to the left of the line, the animal should choose the small reward; below, and to the right, he should choose the large. Imagine that we begin the experiment with t_1 and t_2 chosen to favor the small reward — point A in the figure. Suppose we now progressively (perhaps a step each day) add a constant amount to both t_1 and t_2: This moves the point along the dashed line, parallel to the diagonal, in the figure. It is easy to see that the point must eventually cross the switching line, so that preference should switch from the small to the large reward, as is usually found.

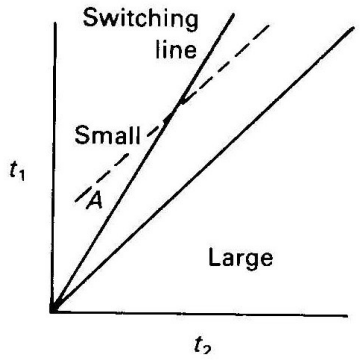

Figure 8.2. Optimal choice for the procedure shown in Figure 8.1. The switching line shows how the animal should weigh delays t_1 and t_2, on the assumption that reward weight over the trial period is maximized. Dashed line shows the effect on predicted choice of adding constant increments to delay.

In another experiment, the two delays, t_1 and t_2, were made equal and the procedure was modified (to a concurrent variable-interval schedule) to ensure that the animals did not prefer one choice exclusively (i.e., the procedure allowed a graded measure of choice)[6]; the effect on preference for the large reward of increasing both delays could then be measured. Increasing both delays, while keeping them equal, is equivalent to movement along the diagonal in Figure 8.2. Because the diagonal diverges from the switching line, such a shift should obviously increase preference for the large reward, and it did (this is a much simplified analysis of the rather complicated, but much used, procedure described in more detail in Note 6).

The strategy I have just described is not as rational as it seems. Equation 8.1 is appropriate only if the animal attends only to trial time, and ignores intertrial interval. So why does it work? The delay experiment can be interpreted in three ways: (a) As a one-time thing: What should the animal do if it is allowed to make the choice between the two delays only once? This is strictly a thought-experiment, of course, because choices are invariably repeated in these experiments — we cannot instruct the animal about the contingencies in any other way. Obviously, the animal should *always* go for the larger reward when offered a one-time choice — unless the delays are so great that they make up a substantial fraction of its expected life span. We would expect no change in preference over the range of delays typically studied, a few seconds or so. Evidently this is not a realistic view. (b) Granted repeated trials, the animal might ignore everything but the trial time. Something like this is tacitly assumed in many discussions of these ex-

periments, but of course it also cannot be correct. Food is necessary for metabolism, and metabolism goes on all the time. An animal that assessed its eating needs only in relation to external circumstances could not long survive. So the third alternative, (c) the animal takes into account the total rate of food delivery, should be correct. How should he take account of both trial and intertrial time?

The simplest way is by a rule that takes account of the overall average food rate given that he makes one choice or the other exclusively. If we denote the ITI duration by T, then from Figure 8.1 the average reward rate, R_A, given that he chooses alternative I exclusively, is obviously

$$R_A = R_1/(T + t_1).$$

Hence the point of indifference between the two choices is

$$R_1/(T + t_1) = R_2/(T + t_2),$$

which can be rewritten as

$$t_1 = (R_1/R_2)t_2 + (R_1/R_2)T\text{-}T.$$

If we rewrite $(R_1/R_2) = w$, for readability, this can be rewritten as

$$t_1 = wt_2 + T(w - 1), \qquad w \geq 1. \tag{8.2}$$

Equation 8.2 is another straight-line switching line, but with an intercept greater than zero on the t_1 axis. We can draw it in the t_1-t_2 space, as shown in Figure 8.3. As before, the region above and to the left of the line denotes exclusive choice of the small reward, R_2, the region below and to the right, exclusive choice of the large reward, R_1. This new switching line has similar properties to the simpler one in Figure 8.2, and explains the effects of increasing t_1 and t_2 in the same way. But it also explains other things.

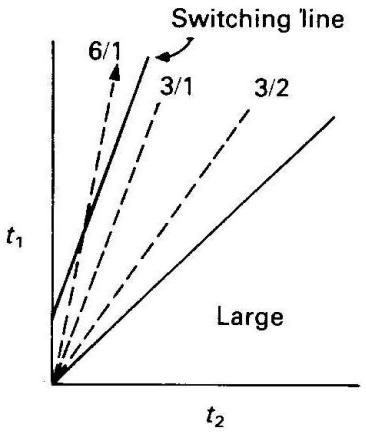

Figure 8.3. Optimal choice for the procedure shown in Figure 8.1, on the assumption that total reward rate is maximized. Reward magnitudes are assumed to be in the ratio $R_2/R_1 = 3$. The dashed line shows the predicted effect on choice of maintaining the delay ratio t_1/t_2 at 6/1, 3/1, or 3/2, while increasing the absolute values of the delays.

The three dashed lines in Figure 8.3 denote experimental manipulations of the ratio of delays, t_1/t_2, tried in a pigeon experiment by Green and Snyderman.[7] They held the ratio of delays constant for each line, but progressively increased the absolute value of the delays and measured the effect on preference. The steepest line is for a delay ratio of $t_1/t_2 = 6$, the next for a ratio of 3, and the shallowest for a ratio of 1.5. The ratio of reward durations (seconds access to a food hopper) was held at 3, which is the slope of the switching line.

Green and Snyderman found that at the 6:1 delay ratio, as the absolute values of the delays increased (in the direction of the arrow), preference shifted toward the smaller reward; whereas a similar absolute-delay increase with the 3:2 ratio produced a preference shift in the opposite direction, toward the larger reward. These shifts correspond to the opposite deviations of the 6:1 and 3:2 delay-ratio lines from the switching line.

My simple model predicts that a change in the absolute value of the 3:1 delay ratio should have no effect on preference, because this line is parallel to the switching line. Green and Snyderman in fact found a shift toward the smaller reward, which suggests that the "real" slope of the switching line is somewhat less than 3:1 — just what we would expect if the more-delayed, large reward is somewhat discounted in value.

Pigeons and rats (and no doubt other species whose choice performance has not been as

well studied) seem beautifully adapted to these choice procedures, sensitively adjusting their preferences in accordance with the overall food rate to be expected from different patterns of choice. Unless special provision is taken to favor nonexclusive choice, the animals' decision rule is simply to choose exclusively, according to the switching line. Discounting of delayed outcomes looks like a limitation, but it makes good adaptive sense — the future is always uncertain, and hunger increases unchecked with time, so delayed food is worth less than present food.

We don't really know how the animals determine the switching line, how they estimate the average food rates to be expected under different strategies. Indeed, it is not at all clear that the switching-line analysis has anything much to do with the mechanisms underlying the behavior. Nor do we know what determines the discount function. The answers to both these questions are obviously related to memory in some way, since the animal's only way to predict the future is by remembering what happened in the past. If the same properties of memory account for other things, such as the limitations on spatial learning or the learning of delayed discriminations, then we will be a little closer to our ultimate goal, which is a limited set of principles that explain everything that an animal can do.

I have so far simplified things by assuming exclusive choice. Most choice experiments are designed to ensure nonexclusive choice, however. I show next how the results of these experiments nevertheless follow the same set of principles.

MATCHING AND MAXIMIZING

In nature, food is rarely distributed evenly through the environment; more commonly it occurs in localized regions of high density — patches. Examples are seeds in pinecones, colonies of insects, a bush of favored browse. An individual meal, the zebra the lion has for lunch, for example, is a patch in this sense. The longer the animal spends looking for food in a patch, the more food he is likely to get, although the *rate* at which he gets it is likely to decrease with time as the patch becomes depleted. The relation between amount of time spent and food obtained is just the *feedback function* (see Chapter 5) for the patch.

Marginal value and momentary maximizing

Suppose that a hungry animal has a fixed amount of time to allocate between two types of depleting patch, "rich" and "lean." This situation is illustrated in Figure 8.4: Cumulative food intake is plotted as a function of time in the patch. Both functions are negatively accelerated because additional food becomes harder and harder to get as food density drops due to depletion; the asymptote (total amount of food in a patch) is higher for the rich patch type. How should an animal allocate its time between these two kinds of patch? What proportion of time should it spend in each, and how often should it switch from one type to the other? Since the feedback functions in Figure 8.4 are negatively accelerated, we know (see Chapter 7) that optimal behavior will be nonexclusive: The animal should spend some time in both types of patch. To maximize overall rate of payoff, the marginal rate of payoff for each alternative should be the same. It is fairly straightforward to show, in addition, that the marginal rate should also equal the average rate of return from the environment as a whole. This result has been termed the *marginal value theorem* (Charnov, 1976).

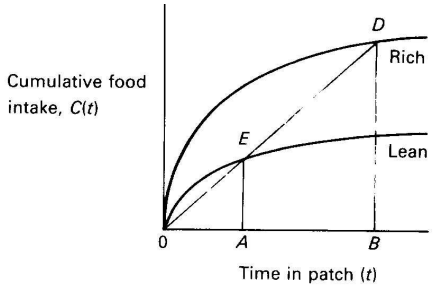

Figure 8.4. Cumulative food intake as a function of time in a patch for "lean" and "rich" depleting patches.

The proportion of time that should be spent in each patch type depends on the form of the depletion (feedback) function. We can get some idea of the proportion by picking a plausible form such as a hyperbola or a power function. Let the cumulative food intake be $C(t)$ for a time t

in a patch. Then a flexible form is the power function

$$C(t) = At^s, \quad 0< s < 1, \tag{8.3}$$

where A and s are constants; if s is less than one, the curve is negatively accelerated. Taking the first derivative to find the marginal value yields

$$dC/dt = Ast^{s-1},$$

which can be rewritten

$$dC/dt = sC(t)/t, \tag{8.4}$$

that is, the derivative of the cumulative food intake function, $C(t)$, can be written as a function of the ratio $C(t)/t$. From the marginal value theorem, dC/dt should be the same for all patches; thus, for two patches 1 and 2, $dC_1/dt_1 = dC_2/dt_2$. If s, the exponent of the depletion function, is the same for both patches, equating the marginal values yields

$$C_1(t_1)/t_1 = C_2(t_2)/t_2,$$

or

$$C_1(t_1)/C_2(t_2) = t_1/t_2, \tag{8.5}$$

that is, the ratio of the times the animal spends in both types of patch is equal to the ratio of the amounts of food obtained; this is termed *matching*. If s is not the same for both patches, then choice proportions should be proportional (not equal) to payoff proportions; this is termed *biased matching*.

This example is an instance of a more general proposition: If animals are acting so as to maximize their rate of payoff, then in situations where the feedback function shows diminishing returns, choice ratios will often match (or be proportional to) payoff ratios.

Concurrent VI-VI

Matching is quite a general finding in operant choice experiments, but the account I have just given does not reflect the actual history of the principle. Matching was discovered in a situation that is the logical complement of the example: In patch foraging, the marginal payoff from the patch the animal is in decreases with time, whereas the payoff for switching to another patch stays constant. In the complementary case, the payoff for staying where you are is constant, but the payoff for switching increases with time. This describes concurrent variable-interval variable-interval schedules, where matching was first proposed as a general principle.[8]

Historically, the concurrent VI-VI procedure had appeal because on a VI schedule the rate of reinforcement varies little with response rate, providing response rate is high enough. Thus (the argument went), any variation in response rate should be a pure measure of the "strengthening" effects of reinforcement, free of the strong feedback interactions between responding and reinforcement characteristic of, for example, ratio schedules.[9] The no-feedback assumption is not wholly accurate, and the reinforcement-as-strengthening theory that underlay the original use of concurrent VI VI finds fewer adherents these days. As we saw in Chapter 5, there is a fairly well defined molar feedback relation even on VI schedules; and as I argued in Chapter 7, this relation appears to underlie many, perhaps all, the distinctive molar properties of interval and ratio schedules.

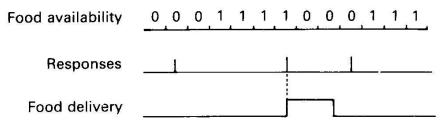

Figure 8.5. Reinforcement contingencies on a random-interval schedule.

The matching law grew out of the belief that reinforcement contributes to something termed "response strength," for which interval schedules offered a convenient measuring rod. The notion of response strength has not been given up (see Chapter 2) — it is a useful concept in the analysis of behavioral competition in discrimination situations, for example (see Chapter 11) — but its significance has changed. I am inclined to look at response strength as a convenient fiction that

provides an intuitive basis for a number of descriptive laws. Here I discuss matching in quite a different way, as something that naturally derives from mechanisms that tend to maximize animals' rate of access to food and other reinforcers.

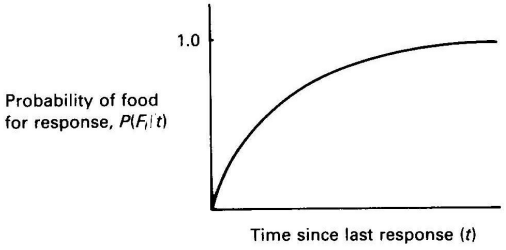

Figure 8.6. Probability of food delivery for a response after time t on a random-interval schedule.

The simplest variable-interval schedule is one in which the availability of reinforcement is determined by a random process (this is also known as a *random-interval* schedule). The way a random-interval schedule operates is illustrated in Figure 8.5, which shows three "time lines." The top line shows time divided up into brief discrete increments, indicated by the short vertical lines. Imagine that during each discrete interval a biased coin is tossed, yielding "heads" (1) with probability p and "tails" (0) with probability 1 - p, where 1 indicates that food is available for the next response. As the brief time increment approaches zero, this procedure increasingly approximates a true *random-interval* schedule. The second time line in the figure shows responses, which can occur at any time. The third line shows food delivery (reinforcement), which occurs immediately after the first response following a 1 (notice that interval schedules are "hold" procedures, in the terminology used earlier since food once "set up" remains set up until a response).

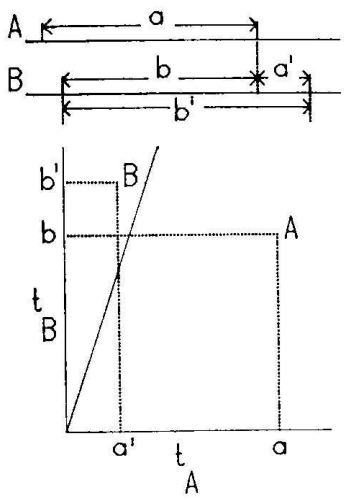

Figure 8.7. Clock-space analysis of the (random) concurrent VI-VI schedule. Event lines at the top show a sequence of A and B choices. The graph at the bottom shows the last A and B choices plotted in a space defined by the times since the preceding A and B choices.

Marginal rate of payoff here is equal to the *probability* that a response will yield food. This probability will obviously increase with time since the previous response, because the longer the time, the more likely a "1" will have occurred — and once a 1 has occurred, food remains available until a response occurs. Obviously, the probability of a "1" is equal to 1 minus the probability that *no* 1 — i.e., only "0s" — occurred during the elapsed time. Let n be the number of time periods since the last response; during each time period, the probability of a 0 is $1 - p$. Each time period is independent, so that the probability that no 1 has occurred during period n is equal to $(1 - p)^n$. The probability that a 1 *has* occurred is just equal to 1 minus this probability, hence the probability of food n time periods after the previous response is given by

$$P(F \mid n) = 1 - (1 - p)^n . \qquad (8.6)$$

Since $1 - p < 1$, as n increases, the quantity to be subtracted from 1 gets smaller and smaller; hence, the probability a response will get food increases the longer it is since the last response. As the discrete time interval becomes vanishingly small, Equation 8.6 is increasingly well approximated by the exponential function, familiar from earlier chapters, so that

$$P(F \mid t) = 1 - \exp(-\lambda t) \qquad (8.7)$$

where t is the time since the last response and λ is a parameter equal to the reciprocal of the mean time between VI "setups," that is, 1/VI value. An example of Equation 8.7 is shown in Figure 8.6.[10]

Suppose an animal is following the hill-climbing rule (momentary maximizing[11]) described earlier; that is, whenever it makes a choice, it picks the alternative with the highest probability of payoff. What form will its behavior take? Clearly the best choice will depend on the

relative times that have elapsed since choice A versus choice B. For example, suppose that A and B choices are reinforced according to (independent) VI schedules of the same value (i.e., the same λ). Then momentary maximizing says the animal should always make the response made least-recently — which implies alternation between the two alternatives. If the two VI schedules are unequal, the rule is just: choose A if $P(F| t_A) > P(F | t_B)$. From Equation 8.7, after some rearrangement, this leads to the simple *switching rule,* choose A iff (if and only if)

$$t_A \rangle t_B(\lambda_B / \lambda_A). \qquad (8.8)$$

To see if choices conform to this rule, it is convenient to represent them in a *clock space* whose axes are the times since the last A and B responses, (i.e., t_A and t_B). Equation 8.8 defines the switching line $t_A = t_B(\lambda_B / \lambda_A)$ in such a space. Figure 8.7 shows two responses to each of two alternatives, and illustrates how they are represented in a clock space. Momentary maximizing requires that all B-responses fall between the switching line and the B axis, all A-responses between the switching line and the A axis. Given an animal well-trained on concurrent VIa VIb (where $\lambda_A = 1/a$ and $\lambda_B = 1/b$) we can plot each response the animal makes in such a space and see how closely its choices conform to this rather stringent rule.

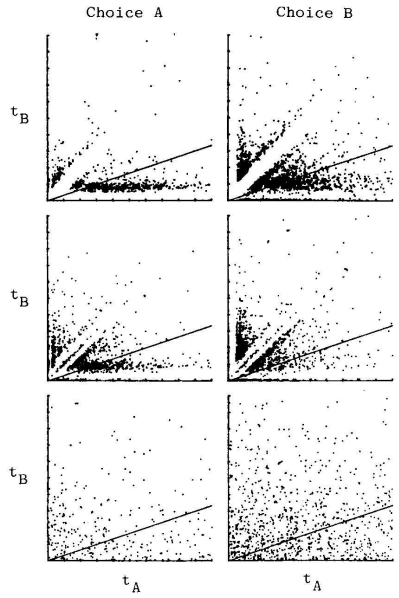

Figure 8.8. Top panels: minority (left panel) and majority (right panel) choice key pecks by a single pigeon during a 1-hr session of a concurrent VI 1 VI 3-min schedule, plotted in the clock space of Figure 8.7. Each dot is a single peck and the axes are the times since the preceding *B* (majority) or *A* (minority) choice. Line through the origin is the momentary-maximizing switching line. During this session, choice and reinforcement proportions matched well. Middle panels: data from the same pigeon during a session when choice and reinforcement proportions did not match well. Bottom panels: simulated choice data, derived from two independent, random processes with response probabilities in 3:1 ratio. Data points are not aligned with respect to the 3:1 switching line (from Hinson & Staddon, 1981).

Figure 8.8 shows such data. Each pair of panels represents a single experimental session and each point represents an individual key peck. So that A and B choices can be distinguished, the left panels just show A (minority) choices, the right panels, B (majority) choices. The top two pairs of panels are data from a single, well-trained pigeon responding for an hour or so on a concurrent VI 3 VI l-min schedule; the switching line is drawn in each panel. The bottom pair of panels show simulated data based on the rule that the animal chooses to respond at random intervals of time (where *random* is defined as in Figure 8.5). There are two differences between the real and simulated data. First, the pigeon shows bands where no responses occur; these depopulated regions correspond to the fixed minimum time it takes for the bird to switch from one response key to another, *switch time.* Second, the pigeon data in the top pair of panels show a strong tendency (confirmed by statistical techniques more refined, if not more compelling, than an eyeball test) for majority choices (B) to cluster above and to the left of the switching line, and minority choices (A) to cluster below and to the right of the line — as the momentary-maximizing hypothesis predicts.

The random data (bottom two panels) fall equally on both sides of the switching line. Although the random probabilities were chosen so that the overall choice proportions would show good matching to payoff proportions, there is no tendency for majority and minority choices to align themselves on either side of the switching line. In other words, obedience to matching does not force conformity to momentary maximizing.

The middle pair of panels show data from the same pigeon as in the top panels, but on a day when he was not matching well (the proportion of minority choices was .30 for the top panel vs. .49 for the bottom; the corresponding payoff proportions were .26 and .29): The alignment of choices with respect to the switching line is poor. It turns out that when pigeons (momentary) maximize well, choice proportions approximately match reinforcement proportions.

The close relation between matching and momentary maximizing is illustrated in Figure 8.9. The figure shows the trajectory traced out in clock space by an imaginary animal that responds every second, and always momentary maximizes. By convention the first response (choice A) occurs at point 1,1. Since the first response is an "A," t_A is instantly reset to zero, as indicated by the leftward dashed arrow, but then both t_A and t_B increase along the 45° line shown (i.e., both times increase at the same rate). After a further second, the representative point is still on the "A" side of the switching line, and another A response occurs; t_A is reset again and the process repeats. By this time, the continued growth of t_B (time since the last B response) has brought the point to the switching line. Our perfect momentary maximizer is now indifferent between A and B, but by convention, we let him make one last A response. The next repeat of the process finds the point on the "B" side of the line, and response B occurs, resetting t_B to zero and leaving the point at 1, 0. Continuation of the process causes the point to track out the triangular trajectory shown in the figure.

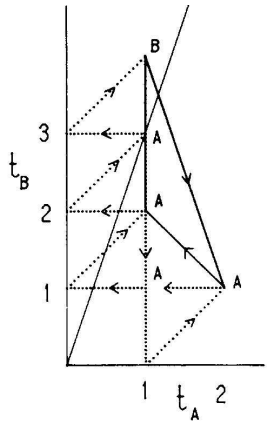

Figure 8.9. Trajectory traced out in clock space when responses are made at 1-sec intervals and choices follow momentary maximizing.

Three things are noteworthy about Figure 8.9: First, momentary maximizing obviously entails nonexclusive choice — no matter where the switching line, exclusive responding to one alternative will cause the representative point eventually to cross the line, forcing a response to the other alternative. Second, the proportion of A's and B's is obviously closely related to the slope of the switching line, hence to the scheduled proportion of payoffs for the two alternatives: the steeper the line, the larger the proportion of A choices entailed by momentary maximizing. Simulations have shown that the commonest result will be *undermatching,* that is, the proportion of responses to an alternative will be somewhat closer to indifference than the proportion of rewards for such responses (e.g., if payoff proportion is 3:1, response proportion will be between 3:1 and 1:1). Third, under most circumstances, the animal should never make more than one response to the minority alternative before switching back to the majority. Choice data conform quite well to these predictions.

Concurrent VI VR

Another procedure that produces nonexclusive choice is concurrent variable-interval variable-*ratio*. This situation can be analyzed in exactly the same way as concurrent VI-VI. A variable-ratio schedule is nothing but a one-arm bandit of the Las Vegas variety: Payoff is strictly proportional to number of the lever presses; *probability* of payoff is constant, equal to one over the mean ratio value, *M*. From Equation 8.6, the condition for switching from the ratio alternative, B, to the interval one is therefore

$$1 - \exp(-\lambda_A t_A) > 1/M,$$

which reduces to

$$t_A > [1n(M/(M - 1))]/\lambda_A, \qquad (8.9)$$

so that the switching line is parallel to the t_B (ratio-schedule) axis.

If animals respond (i.e., make choices) at a steady, high rate, two predictions follow from

this analysis. First, that most responses will be to the ratio alternative, because it takes a time equal to several interresponse times before the probability of payoff on the interval alternative grows (according to Equation 8.6) to a level equal to the constant ratio payoff probability. Second, there will be perfect matching between choice ratios and reinforcement ratios for the two choices. This is easy to demonstrate: If switching to A occurs when $P(F \mid t_A)$ is approximately equal to $P(F \mid t_B)$, then since R_A, the reinforcement rate for A, is equal to $P(F \mid t_A)$ times x_A, the response rate to A (assuming that A choices occur at fixed intervals), and similarly for B, it follows at once that $R_A/x_A = R_B/x_B$, that is, perfect matching. Given reasonable assumptions about overshoot — an A response occurs only after $P(F \mid t_A)$ exceeds the fixed $P(F \mid t_B)$ by some proportion — the expected result is biased matching, favoring the ratio alternative. All these features, a high rate of responding overall, faster responding to the ratio alternative, and biased matching, are characteristic of experimental results with this procedure.

Thus, matching, undermatching (the commonest deviation from simple matching), and biased matching on concurrent VI VR schedules all follow quite simply from momentary maximizing. Active inquiry, and argument, continues, but recent results make it hard to avoid the conclusion that matching on concurrent schedules depends (though perhaps not entirely) on a hill-climbing process.

Overall maximizing

The molar feedback relation between response and reinforcement rates on VI schedules depends on both the average rate and the temporal distribution of responding. Response rate is a single dimension, but the temporal distribution of responses can vary in many ways, and we don't know what constraints to put on temporal variation. Without knowing all the constraints, it is not possible to see whether a particular allocation of responding between VI, or VI and VR, alternatives is the one that maximizes total reinforcement rate. If we also include in the objective function the demands for time of other, non-substitutable activities (see Chapter 7), the position is even less clear. Nevertheless, we can draw some qualitative and even (given suitable simplifying assumptions) quantitative conclusions.

In Chapter 7, I derived nonexclusive allocation of time to different activities from the fundamental economic axiom of diminishing marginal utility (marginal rate of substitution): The more you have of something, the less valuable each additional increment becomes. The decreasing marginal utility here is a reflection of changes in the individual's internal state, a sort of *satiation.* The situations discussed in this chapter have the same diminishing-returns property, except that here it is the relation between a fixed amount of incremental effort expended and the amount of incremental payoff produced that is decreasing. For example, in patch foraging much more food is obtained during the first second spent looking in a patch than during the tenth second, and more during the tenth than the twentieth. Concurrent VI-VI can be thought of in a similar way. Consider the total payoff for an animal responding on a concurrent VI 1 VI 3-min schedule. Suppose that initially the animal spends all its time responding to the VI 1 alternative. Payoff will be about 60/hr if response rate is relatively high. Suppose that the animal now allocates a little time, say 5 min per hour, to the VI 3 alternative. If this time is spread pretty evenly through each hour, the result will be a substantial increase in overall payoff rate, to perhaps 75/hr. A further 5 min spent on the VI 3 will produce a much smaller additional increment, however, and, pretty soon, further increments will produce decreases in overall rate of payoff. A similar argument can be made for concurrent VI VR as the animal allocates more and more time to the VI. At some point in both these situations a single response is likely to produce food with the same probability on each alternative; at this point, marginal values are equal and the animal is maximizing overall payoff. Thus, optimality arguments suggest that the "hold" situations discussed in this chapter, like the allocation of time among non-substitutable behaviors discussed in the last, should yield nonexclusive choice — and for the same reason, diminishing marginal re-

turns.

The range of conditions under which this process will also lead to matching is less well understood — although it is relatively straightforward to find conditions where overall maximizing also implies molar matching.

There is an obvious resemblance between hill climbing and the equation of marginals that is a necessary condition for overall maximization. But as we have already seen, momentary maximizing does not always yield the overall maximum. This is partly because responses occur at finite, not infinitesimal, intervals, and partly because the technique can only find local, not global, maxima. The difference between momentary maximizing and a hypothetical optimal strategy is usually, though not invariably, small.[12] In all cases so far examined, the momentary-maximizing pattern yields overall results closer to matching, and thus to empirical results, than the optimal strategy.

Parameter estimation

An animal must settle three issues if it is to follow a momentary-maximizing strategy: (a) It must decide *when* to make a choice; (b) it must assess the relevant variables (such as post-response times) correctly; and (c) it must in some way assess the relevant schedule parameters (such as scheduled reinforcement rates), so as to correctly estimate the switching line.

The first issue lies outside momentary maximizing itself: The switching-line analysis says nothing about when a choice must be made, only *what* choice should be made, given that a response is about to occur. The second issue can studied directly by looking at how well animals can discriminate time and respond differentially to past events, for example. It turns out that they discriminate time well, but remember individual past events quite poorly — more on this in Chapter 12. The third question is the most difficult. We have no real idea how animals assess the parameters of (for example) variable-interval schedules, although all indications are that they do so with considerable precision.

Parameter estimation can make the difference between strategy that sacrifices short-term gains to long-term losses and one that maximizes overall payoff. For example, in the delay experiments discussed earlier, we saw that pigeons seem to take the intertrial interval into account in their choice; they don't just look at reward size and delay, but also take into account the overall rate of payoff associated with different strategies: Their behavior is better predicted by the switching line in Figure 8.3 than the line in Figure 8.2. The animal seems always to pick the best option available, but its estimate of the best usually takes account of more than just the very next event.

This is true even in concurrent VI-VI schedules. For example, a common modification in free-operant choice is the *changeover delay* (COD); this is a feature that prevents immediate reinforcement for a switch: If the animal's last response was A, any reward set up for choice B can only be collected for a B response at least *t* seconds after the first A-B switch, where *t* is a second or two. The COD means that the probability of reinforcement for a B response after an A (or an A after a B) is zero. An animal that followed momentary maximizing literally would never switch. Imposition of a COD does reduce rate of switching, but not to zero. Hence animals must take account of more than the first post-switch response in estimating the switching line. Most momentary-maximizing errors in concurrent VI-VI take the form of overestimating the majority VI value: Animals seem usually to switch late to the minority choice. It turns out that this bias improves overall maximization: Momentary maximizing based only on the next reinforcement (like the delay switching line in Figure 8.2) doesn't maximize overall reward rate (although the difference is small in the concurrent VI-VI case). Evidently animals are able to incorporate at least some global information into their switching strategies so as make their performance more efficient than a simple one-step-at-a-time momentary-maximizing strategy. We still know very little about how they do this.

SUMMARY

There are two kinds of choice situation: situations that tend to produce exclusive choice, fixation of responding on one alternative or the other; and situations that tend to produce nonexclusive choice, distribution of responses between alternatives. Experiments with the first kind of situation have focused on maximization (picking the best alternative) as the major law of choice; experiments with the second kind of situation have focused on matching (of reward and choice proportions) as the fundamental law of choice.

Situations that assign reward to one alternative or another (single-assignment) and make no new assignment until the reward is collected tend to produce nonexclusive choice, as do *hold* procedures, which keep reward available until the correct response occurs. Situations lacking these features tend to produce exclusive choices. I argue that the same principle — always choose the alternative most likely to pay off (*hill climbing*) — accounts for both exclusive and nonexclusive choice. Hold and single-assignment procedures favor nonexclusive choice because the relative payoff for any alternative increases with time so long as it is not chosen. Such procedures constantly drive the animal away from exclusive choice.

A second theme of this chapter has been the relation between overall (global) maximization and momentary (local) maximization. Animals seem to take each choice as it comes and respond according to a relatively simple decision rule, appropriate to the particular situation. But in every well-studied situation, the decision rule seems sensitively attuned to *global* consequences. Animals do hill-climb, in the sense that they evaluate each choice according to a simple rule that embodies what they know of the probable payoff for each choice. But their rules take into account more than the immediate consequences of a choice. The information for this assessment must come from an animal's past reinforcement history, but we know little of the rules that connect past history to present choice.

NOTES

1. Mathematical learning theory is associated particularly with the names of W. K. Estes (1959; Neimark & Estes, 1967) and Bush and Mosteller (1955) and their students and associates (see reviews in Luce, Bush & Galanter, 1963; and Bower & Hilgard, 1981). The linear learning model briefly described in the text has been ingeniously extended to classical conditioning by Rescorla and Wagner (1972; see notes to Chapter 13). Difference- and differential-equation models for molar properties of free-operant behavior are discussed by Myerson & Miezin (1980) and Staddon (l977a; 1988; Staddon & Horner, 1989).

2. There is now a substantial theoretical and experimental literature on the matching law. The initial experimental paper is Herrnstein (1961); other influential papers are Catania (1963, 1973) and Herrnstein (1970). See de Villiers (1977) and de Villiers and Herrnstein (1976) for reviews of matching as a general law of reinforcement. There are several papers deriving matching from various maximizing principles, e.g., Rachlin (1978) and Staddon and Motheral (1978), although these have not gone unchallenged (Herrnstein & Heyman 1979; Heyman, 1979; Staddon & Motheral, 1979). Reviews are Williams (1988) and Staddon & Cerutti (2003).

3. There is a substantial and rather tangled literature on probability learning and related procedures. In general, most papers are long on experimental results and short on theoretical analysis of the procedures, some of which are quite complicated. For reviews see Sutherland and Mackintosh (1971), Bitterman (1965 1969), and Mackintosh (1969).

Much of the complexity of this literature derives from the way that animals react to spe-

cific physical apparatus. For example, rats in T-mazes are much more inclined to attend to the place where they find food than to other stimulus aspects, such as the pattern of the goal box. Consequently, in experiments where position is irrelevant, the rats often respond to position anyway, either by going back to the side where they were just fed ("reward-following") or, more commonly, *avoiding* that side (*spontaneous alternation*). Another common pattern is to choose the same side repeatedly, even when position is irrelevant (this is termed a *position habit*). These built-in response tendencies obscure patterns in the choice sequence specifically adapted to the choice procedure. The discussion in this chapter is of the ideal case, where these confusing and largely irrelevant patterns have been eliminated. This ideal is approximated by very well-trained animals working in choice situations with highly salient (preferably spatially separated) alternatives. Response patterns such as alternation and reward-following are integral to the acquisition of adaptive strategies, however, and will be discussed in more detail in later chapters.

4. For an extended theoretical discussion of optimal choice on a variety of simple discrete-trial and free-operant choice procedures, see Houston and McNamara (1981) and Staddon, Hinson, and Kram (1981).

5. See, for example, Rachlin (1970), Rachlin and Green (1972), and Herrnstein (1981).

6. Navarick and Fantino (1976). Nonexclusive choice in these procedures is ensured by making both choices continuously available on an interval schedule (rather than presenting them after a fixed ITI). The procedure is as follows: A hungry pigeon is confronted with two response keys, both lit with white light. Independent, equal VI schedules (mean value $2T$) operate on both keys so that the average time to reward on both, taken together, is T. When a VI sets up, a peck on the key changes the color on that key, and turns off the other key. Food is then delivered after a delay t_i. After the animal eats, the two keys are again lit white, and the cycle begins again. This is known as a *concurrent chain* schedule (in this case, concurrent VI-FT, where the VI operates in the first *link* and the fixed-time schedule in the second link). Numerous changes are rung on this basic theme. For reasons discussed later in the chapter, a concurrent interval schedule produces nonexclusive choice. Hence, given two equal first-link VIs, and other things being equal, the animals will tend to more or less alternate between the two alternatives. A shift in preference, because of changes in delays in the second links, shows up as a bias in this preference.

The concurrent-chain procedure is more complicated than the discrete-trial procedure described in the text and more difficult to analyze exactly in optimality terms. The *changes* in preference to be expected from changes in delays can be derived in the simple way I describe, however, even though exact predictions of response proportions probably require something more elaborate.

7. Green and Snyderman (1980). Their experiment, like the experiment of Navarick and Fantino (1976), used concurrent-chain VI FI schedules.

8. The concurrent VI-VI situation can be represented as in Figure 8.4 if we plot on the horizontal axis the proportion of responding (out of a fixed total) devoted to one or other schedule and on the vertical axis the rate of reinforcement associated with the proportion of responses. It is easy to see that for each choice, the rate of reinforcement will rise with the proportion of responses made to that choice, up to an asymptote determined by the VI value. If this molar feedback function (see Chapter 5) has the property that its derivative can be expressed as the ratio of reinforcements obtained to responses made, $R(x)/x$ (using the symbols of Chapter 7), then allocating responses so as to maximize total reinforcement will also lead to matching (see Staddon & Motheral, 1978).

9. Ratio schedules correspond to *non*-depleting patches: If "foraging rate" (lever-press rate) is constant, cumulative payoff increases linearly with time. The marginal-value theorem (not to mention common sense) therefore predicts that animals should concentrate exclusively on the richer patch (i.e., the one with the steeper cumulative-gain function). Under most circumstances, they do (e.g., Herrnstein & Loveland, 1975).

10. The curve in Figure 7 has the same negatively accelerated form as the patch-depletion function in Figure 5, but its significance is quite different. *Probability* of reinforcement for a response is equivalent to *marginal* cumulative food intake, that is, to *rate* of food intake. Imagine a response occurring at a fixed interresponse time t with probability q that each response will be reinforced. The expected time between reinforcements, E, is equal to t times q, the probability that a response will be reinforced, plus 1 - q times the expected time plus t, that is, a recursive equation of the form, $E = tq + (E + t)(1 - q)$, which reduces to $E = t/q$. Thus, reinforcement rate is equal to q/t: For a fixed response rate, reinforcement rate is proportional to probability of reinforcement for a response.

11. Momentary maximizing was first suggested as an explanation for matching by Charles Shimp (1966, 1969). His initial theoretical analysis of a rather tricky problem was incomplete, and testing the hypothesis appeared complicated, so that his arguments at first failed to persuade many who were more taken with the empirical simplicities of molar matching. Subsequently, interest in this view revived. The simpler theoretical analysis described in the text has appeared and new experiments supporting momentary maximizing as a substantial component, at least, of operant choice have been published (Silberberg, Hamilton, Ziriax, & Casey 1978; Staddon l980b; Staddon, Hinson, & Kram, 1981; Hinson & Staddon, 1983) - although dissenting views are not lacking (e.g., Nevin, 1979; de Villiers, 1977).

12. Momentary maximizing may fail to maximize overall payoff even in very simple situations. For example, in a discrete-trial procedure where food is assigned to one or other choice, but is then available for a response only with some probability r, momentary maximizing yields the optimal strategy only when food is not held, that is, when probability r is sampled on every trial. Momentary maximizing is not optimal in concurrent VI-VI or its discrete-trial equivalent. In all these cases, the shortfall is small, however (Staddon, Hinson, & Kram, 1981).

The problem with globally optimal strategies in these situations is that they place a substantial load on memory. For example, in the concurrent VI-VI situation with nonrandom VIs, given that the animal has a fixed number of responses to "spend" in a fixed session time, the optimal strategy consists of a *trajectory* through the clock space in Figure 8.7. (Equating marginals would mean placing each choice on the switching line, but the geometry of the situation makes this impossible in most cases.) In order to follow such a trajectory, the animal might need to be guided by several of its past choices and their times of occurrence — a formidable load on memory. The great virtue of hill climbing is that it nearly always approximates the global maximum, while placing minimal demands on memory. In the concurrent VI-VI case, the animal need only remember the times that have elapsed since the *last* A and B choices.