Psychological Review

VOLUME 89 NUMBER 3 MAY 1982

Matching, Maximizing, and the Hyperbolic Reinforcement Feedback Function

Dražen Prelec
Harvard University

A theory of hyperbolic feedback functions for schedules of reinforcement is developed, followed by an analysis of matching and maximizing behavior in an environment characterized by such feedback functions. The hyperbolic function classifies schedules along two dimensions: one that measures the time and one that measures the responses that are needed to collect a unit of reinforcement. Among other results it is shown that (a) both response rules predict *pairwise linearity*, a condition which states that absolute rates of response on any two schedules are mutually constrained by a linear function, (b) matching and maximizing rules predict identical behavior if and only if the predictions of either one are consistent with Luce's choice axiom, and (c) the hyperbolic feedback function is preserved under aggregation of response classes. The evidence collected from single and concurrent schedules of reinforcement strongly favors the matching interpretation of equilibrium behavior, as subjects do not seem to be influenced by the marginal trade-offs that define the maximizing behavior distribution. It is argued that insensitivity to marginal variables undermines not only the specific hypothesis of reinforcement-rate maximization but also the more general theories of value maximization developed by Rachlin, Staddon, and others.

I try to accomplish several things in this article. Taking the variable-ratio (VR) and variable-interval (VI) schedules as basic elements, I construct a general representation of schedules in which reinforcement depends conjointly on effort and on the passage of time. Second, I derive in detail the implications of matching and maximizing for behavior on free-operant schedules and compare these implications with available data. A large part of this essay is thus a contribution to the recent controversy between the matching and maximizing approaches to continuous choice behavior (Herrnstein & Heyman, 1979; Heyman, 1979; Heyman & Luce 1979a, 1979b; Rachlin, 1979; Staddon & Motheral, 1978). Finally, I suggest that understanding behavior maintained by schedules of reinforcement—especially those schedules, such as the VI, that impose hidden trade-offs between present and future reward—is a prerequisite for understanding the more general problem of behavior allocation (Rachlin & Burkhard, 1978; Staddon, 1979) and, furthermore, that only by means of a satisfactory quantitative theory of operant performance can one begin the reconstruction, this time on solid empirical foundations, of the economic theory of value

---

Preparation of this paper was supported by National Institute of Mental Health Grant MH-15494 (R. J. Herrnstein, principal investigator) and National Science Foundation Grant BNS78-10685 (R. D. Luce, principal investigator) to Harvard University. I would like to express my appreciation and thanks to my teachers and advisors, to Richard J. Herrnstein for the help he offered in this endeavor and for his assistance and encouragement throughout the years, and to R. Duncan Luee for his initial reading of the paper followed by invaluable suggestions and comments.

Requests for reprints should be sent to Dražen Prelec, William James Hall, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138.

0033-295X/82/8903-0189$00.75

and substitutability (Rachlin & Burkhard, 1978; Rachlin, Kagel, & Battalio, 1980).

By way of introduction consider the following experiment: On a discrete trial procedure, the subject repeatedly makes a choice between two alternatives, A and B; occasionally his or her choice is rewarded with money. The probability of reward for selecting A is constant, say .03, and independent of any other event. The probability of reward for selecting B obeys a more complex rule: On each trial, whether the subject chooses B or not, there is a certain probability, say .02, that reward will "set up"; once the reward is set up, the first selection of B is sufficient to collect it. From the subject's perspective the probability of reward for B increases with the number of trials that B was not chosen. If B was last chosen two, three, or four trials before, then the probability of reward for B rises from the initial .02 to .0390, .0588, .0776, and so forth, respectively.

This miniature experiment is interesting for several reasons. First, the distinction between rational and irrational behavior, which is so easily drawn in more elementary choice situations, cannot be made here with the same conviction. There is more than one defensible response rule. In fact, we can identify three: (a) global maximization, the rule that ensures the maximum overall frequency of reward; (b) momentary maximizing, the rule that always selects the alternative that has the greater probability of reward; (c) matching, the rule that equalizes the long-run conditional probability of reward on the two alternatives. All three rules have some initial plausibility; not one could be dismissed as an irrational solution to the problem of what the subject should do; a fortiori, not one may be dismissed as an impossible solution to the problem of what a subject will do.

A purely formal question of interest is whether these three rules really conflict or merely reflect different definitions of a single response strategy. For this miniature experiment one can quickly show that the three rules identify three genuinely different modes of responding. Conformity to any one rule will necessarily imply a violation of the other two. Global maximization requires responding in cycles of twelve selections of A followed by one selection of B; the overall reward rate is the highest—4.55 rewards per 100 trials—but the average conditional probability of reward when B is selected is almost eight times greater than when A is selected. Momentary maximization requires strict alternation between A and B. The overall reward rate is 3.48, and the average conditional probability of reward when B is selected is one-third greater than when A is selected. Last, although matching does not uniquely specify any one response pattern, it does specify that the conditional probabilities of reward for A and B be equal, and consequently, the overall rate of reward is fixed at 3.00—the lowest overall rate dictated by the three rules.

This experiment captures a feature common to many day-to-day decisions. To see this one can recast the experiment in a context familiar to choice theorists: the menu problem. Imagine yourself going to the same restaurant for lunch every day. The menu has only two entries and they do not change: an ordinary dish A, whose appeal, though not great, remains steady, and a more exotic dish B, which must be sampled infrequently to be fully enjoyed. The dilemma, qualitatively at least, is of the same nature as the one in the original experiment. Dish A has the same value every day, regardless of past history, whereas the value of dish B increases in some fashion with the number of days it has not been eaten. The three rules are (a) maximize the overall hedonic potential of the meals, (b) always select the more appealing dish, and (c) choose so as to make the dishes equally enjoyable in the long run.

The study of these problems has gone on in psychology for some time now, not in the context of traditional choice theory, but in work with schedules of reinforcement. Alternative A, in this experiment, was a variable-ratio schedule—it reinforced each response with a fixed probability; alternative B, on the other hand, was essentially a variable-interval schedule—it required a response and the passage of a certain number of trials for reinforcement to occur. The data generated by this class of operant procedures is therefore a body of available evidence against which the claims of different response rules can be assessed; to carry out such an assessment is one of the main goals of this article.

Background

The matching and maximizing theories I will be concerned with are designed to account for equilibrium behavior in operant-conditioning experiments. In a typical procedure a deprived subject has the opportunity to respond on one or more schedules of reinforcement. The experimenter fixes the parameters of a schedule and waits until behavior is in equilibrium, that is, until it reaches some predetermined degree of stability, for example, from session to session or from hour to hour. The experimenter then changes the parameter values and notes the resulting change in the equilibrium point. The fundamental datum, then, is the relation between the schedule parameters and the steady-state, asymptotic behavior distribution of the subject. This relation is what theory has to explain.

A major impetus for a quantitative analysis of steady-state responding came from the early work on matching behavior in concurrent variable-interval schedules (Herrnstein, 1961). In Herrnstein's experiment, a hungry pigeon could peck at two keys, and occasionally a peck on one or the other key would be followed by food. Access to food was controlled by two VI schedules of reinforcement, that is, schedules that rewarded the first peck after the passage of a variable time period. As the parameters that determined how frequently the schedules would make food available were varied, the following relation between pecks and reinforcements emerged:

$$\frac{b_1}{b_1 + b_2} = \frac{r_1}{r_1 + r_2}, \qquad (1)$$

where b_1 and b_2 are the average rates of responding on the two keys and r_1 and r_2 are the average rates of reinforcement received from the two keys. (The averages are computed over total session time.) This relation has been called the *matching* relation or the *matching law*, and it has been found to hold for a wide variety of species and in a wide variety of situations (de Villiers, 1977; Herrnstein, 1970).

Subsequently, several authors attempted to derive the matching relation from optimizing principles. They did this in essentially two ways. The first approach, started by Shimp (1969), made use of the notion of momentary maximizing, which postulates that the variables controlling the subject's behavior are the local probabilities of reinforcement on the two schedules. A subject who followed the rule of momentary maximizing would, then, respond on one schedule as long as the probability of reinforcement was higher on that schedule than on any other. Since 1969 there has been considerable debate whether this is a true description of the subject's behavior on the molecular level.

The second approach, developed by Rachlin and his associates (Rachlin, Green, Kagel, & Battalio, 1976), was based on the idea that subjects behave as if they are maximizing the overall or global rate of reinforcement, that is, the rate of reinforcement averaged over the experimental session. More recent elaborations of this model have allowed for the possibility that the subject is maximizing some utility or value function that includes among its arguments the time spent resting and the time spent responding (Rachlin & Burkhard, 1978; Staddon, 1979).

To summarize, a large body of data on equilibrium behavior in operant experiments requires explanation, and the current theoretical opinion about it is divided along two lines. One point of dispute is about the proper level of analysis: Is it the molecular level or is it the molar, or aggregate, level? The other point of dispute is whether the subject's behavior is best described by a process of matching response ratios to reinforcement ratios or by some sort of reinforcement-maximization process.

In constructing and comparing two molar models, one based on matching and one on maximizing principles, I attempt mainly to answer the second question. On the first question my position is essentially agnostic: I see research on the two levels as being complementary rather than competitive. A successful molar theory attains greater generality and simplicity at a price; it usually says nothing about the molecular structure of responding. At the same time, such a theory will always restrict the range of acceptable molecular theories and thus accelerate progress on that level as well.

Response Rules and Feedback Functions

To understand the differences between matching and maximizing, it is important first to know the ideas they share, and foremost among these is the notion that behavior can be usefully analyzed as part of a feedback system. According to Baum (1973, 1981), who has formulated this idea with particular clarity, there are two parts to the organism–environment feedback system: the control that the organism exerts over the environment and the control that the environment exerts over the organism. The extent to which the organism can control its environment is specified by the schedule of reinforcement. In all its aspects the schedule is, in principle, known to the experimenter. On the other hand, the manner by which the environment controls behavior is not known by the experimenter; that is the element of the feedback system that must be uncovered. Let us consider how the experimenter might go about doing that.

For any given set of schedule-parameter values, the experimenter records a description of the subject's behavior paired with a description of the environment. The two types of control just mentioned might be thought of as two independent restrictions on (and hence subsets of) the set of all possible joint descriptions of behavior and environment. The description, mathematical or otherwise, of the restriction imposed by the schedule is called a *reinforcement feedback function* (RFF; Baum, 1973; Prelec & Herrnstein, 1978). The other restriction, corresponding to Skinner's "functional relations" (Skinner, 1938) and Baum's "O-rules" (Baum, 1973), will be called the organism's *response rule.*

It follows from this formulation of the problem that at any time, the description of the behavior and the environment must violate neither the restriction imposed by the RFF nor the restriction imposed by the response rule. A subsequent change in the schedule parameters alters the shape of the RFF, thus displacing the locus of its intersection with the function describing the response rule. By carefully changing the RFF and observing the resulting changes in behavior, the experimenter, in effect, monitors the changing intersection of a known set with an unknown one, thus coming to an understanding of the shape of the unknown set—the set identical with the response rule he or she has chosen to study.

It is here that matching and maximizing yield divergent claims. This can be clearly seen even in the simple concurrent-schedule experiment. Both theories initially implicate three variables: the rates of responding on the two operands (b_1 and b_2) and the overall rate of reinforcement (r). The initial feedback function (f) is

$$r = f(b_1, b_2). \tag{2}$$

This function shows that the rate of reinforcement depends, to a first approximation, on the two rates of responding. An adequate response rule cannot, however, be constructed from just these three variables. The rate of reinforcement may determine in some way the total rate of responding, but it implies nothing about how responding should be divided between the two operands. It is well known that in concurrent-schedule experiments a given total reinforcement rate is compatible with many different response-rate distributions. Equation 2, therefore, is not an adequate feedback function, although it is a good starting point for one. Some more information is needed.

The matching equations identify one possible solution to this problem, and that is to partition instances of reinforcement into two categories: those that were immediately preceded by a response on one operand and those that were immediately preceded by a response on the other. The total rate of reinforcement is divided into two components, r_1 and r_2, that sum up to the total rate r. The feedback function, then, has to describe how r_1 and r_2 depend on the rates of responding:

$$r_1 = f_1(b_1, b_2),$$

$$r_2 = f_2(b_1, b_2).$$

The matching equation implies, of course, much more than the dimensions of the controlling environmental variables; it implies a certain mathematical relationship—matching—between these variables and response rates. This mathematical relationship should not obscure, however, the significance of this particular choice of environmental variables.

Maximizing, on the other hand, involves a different set of controlling variables. To a subject who wants to maximize some function of reinforcement rate and response rate, the crucial variables are the marginal changes in reinforcement rate with respect to marginal changes in responding, that is, the partial derivatives of r, as given by Equation 2, with respect to b_1 and b_2. The appropriate feedback functions are the ones that show how these marginal rates of reinforcement ($\partial r/\partial b_i$) depend on the rate of responding:

$$\frac{\partial r}{\partial b_1} = \frac{\partial f}{\partial b_1}(b_1, b_2)$$

$$\frac{\partial r}{\partial b_2} = \frac{\partial f}{\partial b_2}(b_1, b_2).$$

The two theories thus differ on a deeper level than the level of response-rule specification. Fundamentally, they are in dispute over the nature of the terms in which the response rule and the feedback functions should be defined. A matching response rule is only one of a large family of possible response rules (e.g., Baum's, 1974, generalized matching law) that are based on rates of reinforcement. Likewise, maximizing is only one of many possible response rules based on marginal trade-offs. It should be clear, however, that any response rule that depends materially on variables other than the partial derivatives of the feedback function, f, is a priori not an optimal strategy. In particular, any response rule that distinguishes reinforcements that follow responses on one operand from those that follow responses on the other will lead to nonoptimal performance. Subjects who distinguish these two kinds of reinforcements are behaving superstitiously, that is, they are taking into account information that they should ignore.

The underlying question I shall attempt to answer, then, is which, if either, of the two sets of variables is faithful to the animal's natural classification of its behavior and environment in the operant experiment. It is possible that both sets play a role; however, by making the opposition between them as clear as possible, I try to see which one is of primary importance.

Assumptions of the Special Theories

The immediate goal is to account for the results of a small group of experiments—experiments in which the subject responds on a VI or VR schedule of reinforcement or perhaps some combination of them. However, this task becomes theoretically simpler if we analyze a more general experiment, one that will contain as special cases all the actual experiments that we are interested in. The standard concurrent-schedule experiment will be generalized in two ways. First, the number of schedules will not be limited to two but may be arbitrarily large. Second, each schedule will be a member of a general class of schedules with hyperbolic RFFs. This class, which will be precisely defined later on, includes the VR and the VI, but it also contains many other schedules.

Measurement of Responding

I assume that the subject's behavior can be broken down into n response classes, hereafter called *activities*. Responses belonging to a small number (m) of these activities are reinforced by real schedules of reinforcement, whereas responses belonging to the remaining $n - m$ activities occur spontaneously. A complete description of the subject's behavior, so far as it concerns us, is a list of n frequencies, $b_1, b_2, \ldots, b_n$, that specify how often responses belonging to the n activities occur per unit of time. In practice, attempts to measure and report the frequencies of spontaneously occurring responses are rare, which means that we have to be content with the incomplete list, $b_1, b_2, \ldots, b_m$. For this reason, also, I will sometimes refer to spontaneous activities as unobserved or background activities.

To qualify as an activity, a response class has to satisfy several requirements. First, it must be an operant class: It should be possible to reinforce and punish responses of the class by traditional techniques of operant conditioning. Second, responses belonging to that class, say class i, must be able to occur at any frequency between zero and a certain maximum frequency k_i, specific to that class, without undergoing appreciable changes in response topography. This second condition may never be satisfied exactly, but it will serve to distinguish those responses that can be executed naturally at any frequency within their dynamic range from those that either always occur in the neighborhood of some favorite frequency or show, at different frequencies, drastic changes in topography.

Third, responses should occur randomly in time or, at least, approximately so. By this requirement I wish to exclude from consideration responding maintained by fixed-interval or fixed-ratio schedules.

Instead of reporting response frequencies, researchers sometimes report proportions of session time spent responding. My definition of an activity is entirely compatible with time-based measures of responding, although for the sake of simplicity I will always refer to frequencies of response. Time measures can easily be converted to frequency measures by defining the discrete response to be a fixed tiny amount of time spent responding.

How should one decide which response measure to use when more than one is available, for example, when the experimental apparatus records both discrete responses and the time spent responding? If the two measures (b' and b'') are similarity transformations of each other ($b' = cb''$, for some constant c), then either one will do equally well. If, on the other hand, they are not so simply related, then either one or the other should be discarded. There are two possible criteria for making a choice of this kind: The first is to see which measure yields more orderly data, and the second is to see which measure points at a behavioral unit that is more uniform topographically at different response frequencies. One could, for example, take photographs of the subject at the moment when the two recording apparatuses register a response and then, collecting photographs taken from sessions in which the overall frequency of response was different, make a judgment as to which of the two sets of photographs define a more homogeneous class.

The fourth requirement that activities must satisfy is stated formally by Assumption 1.

ASSUMPTION 1. *Response frequencies obey a linear constraint*:

$$\sum_{i=1}^{n} b_i/k_i = 1.$$

The total amount of behavioral resources available for operant responding is constant, and the fractions b_i/k_i show how this total is allocated among the n activities. Because each b_i can range from zero to k_i, it follows also that each activity can absorb none, all, or any intermediate amount of the total.

Assumption 1 should be carefully distinguished from assumptions about purely temporal constraints on responding that have been a part of several recent theories of operant behavior (Baum & Rachlin, 1969; Rachlin & Burkhard, 1978). The reader may wonder why, if there is going to be an explicit response constraint in the theory, I did not choose to define responding in time units and thus have the constraint equation hold automatically. There are several reasons. First, time measures are derived from a series of discrete transition events (events which indicate that one activity has stopped and another has begun) whose definition is as arbitrary and uncertain as the definition of a standard operant. The difficulties involved in defining an activity correctly cannot be removed by excluding, on theoretical grounds, a variety of otherwise acceptable measures of responding.

Second, time measures have the dubious advantage of simultaneously providing two sets of values: the durations of time spent in each activity (b_i) and the relative response shares (b_i/k_i). When dealing with activities for which the notion of response duration is particularly ambiguous, a technical decision about how many seconds must pass before a given response timer starts or stops will affect not just the raw datum b_i but also the supposedly invariant quantity b_i/k_i. Response-rate measures do not share this difficulty because they make us treat the practical problem of defining an accurate and sensitive response measure as entirely separate from the theoretical and statistical problem of estimating the associated maximum rate k_i. A different definition of what constitutes a discrete response should not drastically change the estimate of b_i/k_i because the change in the measured value of b_i will be canceled by a roughly proportional change in the estimated value of k_i.

Third, time measures force us to adopt an unnecessarily specific interpretation of the response constraint (Assumption 1), namely, that the total amount of behavior is constant because the subject can do only one thing at a time. For all we know the subject can engage in more than one activity simultaneously. In my theory the linearity of the response constraint equation is assumed to be only a first approximation to a more general constraint equation: $G(b_1, b_2, \ldots, b_n) = 1$. In a theory committed to time measures, the linearity of the response constraint is a permanent fixture.

The next two assumptions are largely technical.

ASSUMPTION 2. *The total number of activities (n) is not less than two; the total number of observed activities (m) is not less than one.*

This assumption merely states that there is at least one real schedule in the experiment ($m \geq 1$) and that it does not control all behavior ($n \geq 2$); otherwise, there would either be nothing to observe ($m = 0$) or nothing to explain ($n = m = 1$).

ASSUMPTION 3. *Frequencies of responding are greater than zero for all n activities*:

$$b_i > 0, \quad i = 1, \ldots, n.$$

The theory is not concerned with all the responses that the subject could execute, but only with those that are actually occurring in the experiment. Behaviors that do not occur have been discarded from the analysis.

Measurement of Reinforcement

A reinforcing event is a brief change in the environment of a fixed duration. To say than an observed activity is reinforced means that some responses belonging to that activity are immediately followed by a reinforcing event. The rate of reinforcement for that activity is the rate at which such events occur during the experimental session. The duration of the reinforcing events is ordinarily not considered a part of total session time. Because unobserved activities, by assumption, conform to the same behavioral laws as activities reinforced by real schedules of reinforcement, I must assign to each unobserved activity having a postulated nonzero response rate some source of reinforcement, which, for the sake of a uniform vocabulary, I will call a *schedule of reinforcement*. Reinforcement delivered by such a schedule cannot, of course, be observed, but it could, in principle, be estimated in a variety of ways. One could try to find out, perhaps, how much punishment is required to suppress the activity completely or how responding on other, concurrently available schedules is affected when this activity is mechanically prevented.

Physically different reinforcing events are not interchangeable; they affect behavior differently, and this gives rise to a problem of scaling. I will assume that all distinct reinforcing events used in an experiment can be mapped onto a ratio scale, called a *value scale*, and that the values assigned by this mapping will not be influenced by the frequencies of obtained reinforcement. In the context of matching theory, the value scale ought to be interpreted as a measure of the reinforcing power of physically different reinforcing events, whereas in the context of maximization theory, this same scale should be interpreted as the marginal utility of different event commodities (marginal utilities are, indeed, unique up to multiplication by a positive constant). The dimensions of reinforcing power and marginal utility play identical roles within their respective theories, and it is only by virtue of this mathematical symmetry (which will be derived shortly) that we are able to speak of a single scale but actually refer to two.

The Hyperbolic Reinforcement Feedback Function

The RFF, which is a function h for any particular schedule of reinforcement, describes how response rates and reinforcement rates are constrained to covary on that schedule: $r = h(b)$. It is true that on some schedules, such as the variable-interval schedule, the response rate by itself does not uniquely determine the reinforcement rate; the RFF for such schedules should really be written as $r = \varphi(b, \mathbf{z})$, where $\mathbf{z}$ is a catchall vector variable representing all the molecular response statistics (interresponse time distributions and such) that can have an independent effect on obtained reinforcement rate. By restricting myself to the molar variable, b, I am accepting as true a substantive empirical hypothesis about the orderliness in the subject's behavior, which is that for any schedule and any individual organism the relevant molecular statistics are themselves determined within reasonable limits by the overall rate of response: $\mathbf{z} = \mathbf{z}(b)$. This assumption then justifies the reduction of $\varphi[b, \mathbf{z}(b)]$ to $h(b)$. The molecular properties of responding may still appear in h but only in the form of constants specific to a particular organism–schedule combination.

Schedules of reinforcement vary in the extent to which they make the passage of time and the execution of responses a necessary condition for reinforcement delivery. What I would like to do is formalize at the level of the feedback function what it means for a schedule to impose a certain time requirement and a certain response requirement on reinforcement.

I will refer to the average interreinforcement interval as the cost of reinforcement in time units and to the average number of responses made during one interreinforcement interval as the cost of reinforcement in response units. Consider an RFF having a large number of different parameters $p_1, p_2, \ldots p_n$. A parameter p will be called a *time requirement* if changing the value of p by Δp will, ceteris paribus, change the cost of reinforcement in time units by Δp. Likewise, a parameter p will be called a *response requirement* if changing the value of p by Δp will, ceteris paribus, change the cost of reinforcement in response units by Δp. The ceteris paribus clause covers, in both cases, the values of the remaining parameters and response-rate level.

The scope of the theory is now restricted to those schedules that fall within the two-dimensional family, defined by the time and response requirement parameters.

ASSUMPTION 4. *The reinforcement feedback function for any schedule under consideration, h_i, is determined uniquely by the values of two parameters, a time requirement, d, and a response requirement, a:*

$$r_i = h_i(b_i) = H(d_i, a_i, b_i).$$

The defining function H is also continuous in d, a, and b.

It follows from a simple functional equation argument that the only functions satisfying Assumption 4 are the rectangular hyperbolas:[1]

$$r_i = \frac{b_i}{d_i b_i + a_i}. \tag{3}$$

Taking reciprocals of Equation 3 shows that the postulated relationship between the average interreinforcement interval and the time requirement does in fact obtain:

$$1/r_i = d_i + a_i(1/b_i). \tag{4}$$

---

[1] By definition of the time and response requirements we have

$$1/r = 1/H(d, a, b) = H_a(a, b) + d, \tag{1'}$$

$$b/r = b/H(d, a, b) = H_d(d, b) + a. \tag{2'}$$

It is also implicit in the definition that the time requirement is measured in units of time and the response requirement in units of response number. The arbitrariness of the unit for time (seconds, hours, etc.) then yields one functional equation,

$$H(\theta d, a, b/\theta) = r/\theta = (1/\theta)H(d, a, b) \quad (\theta > 0), \tag{3'}$$

and the arbitrariness of the accounting unit for responses (ones, dozens, hundreds) produces another one,

$$H(d, \theta a, \theta b) = r = H(d, a, b) \quad (\theta > 0). \tag{4'}$$

From Equations 1′ and 2′ it follows that

$$H_d(d, b) = bH_a(a, b) + db - a,$$

implying that $[bH_a(a, b) - a]$ is a function of b only:

$$bH_a(a, b) - a = \mathrm{G}(b).$$

Substituting back into Equation 1′ gives

$$r = \frac{b}{\mathrm{G}(b) + a + db}.$$

Applying Equation 3′ shows that G is a constant function: $\mathrm{G}(b/\theta) = \mathrm{G}(b) = \mathrm{G}_c$. Applying Equation 4′ shows that $\mathrm{G}_c/\theta = \mathrm{G}_c$ or that the constant G_c is equal to zero.

Multiplying both sides of Equation 4 by b then shows that the postulated relationship between the average number of responses expended per single reinforcer and the response requirement also holds:

$$b_i/r_i = d_i b_i + a_i\,.$$

Equation 4 suggests that it is possible to conceptually divide the average interreinforcement interval into two components, a constant delay equal to the time requirement, d, and a variable delay, a/b, which can be interpreted as the concatenation of the a interresponse times (IRT) needed to complete the response requirement.

Sometimes it will be more useful to write the hyperbola using the reciprocals of the parameters d and a. If we let $s = 1/d$ and $p = 1/a$, then Equation 4 may be rewritten as

$$r = \frac{sb}{b + sa} \tag{5}$$

or

$$r = \frac{spb}{pb + s}\,.$$

The parameter s, being the reciprocal of the minimum time requirement, is the maximum rate of reinforcement that the schedule will deliver, whereas p, being the reciprocal of the minimum response requirement, is the maximum probability that any one response will be reinforced. Equation 5 is probably the most familiar form of the hyperbolic function; it shows directly how the rate of reinforcement reaches an asymptote at the maximal rate as the rate of responding increases indefinitely.

To get a better understanding of how the hyperbolic RFF depends on the parameter values, let us restrict our attention to a small subset of all possible parameter combinations. In particular, let us focus just on those combinations of d and a that produce a function that goes through one point, let us say the point where the response rate is $b^{\max}$ and the reinforcement rate is $r^{\max}$. We can think of $b^{\max}$ as the maximum rate of response and $r^{\max}$ as the corresponding rate of reinforcement. The response requirement, a, now indexes different RFFs, all of which pass through the points (0, 0) and ($b^{\max}$, $r^{\max}$). For the present purposes they may be classified into five distinct categories (see Figure 1):

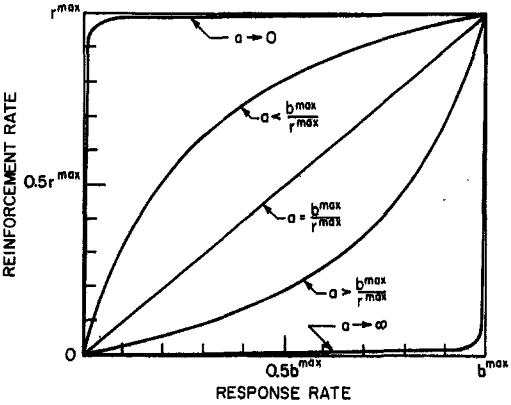

Figure 1. The hyperbolic reinforcement feedback function, indexed by five different values of the response requirement parameter (a).

1. If $a = 0$, then the RFF degenerates into the constant function: $r = r^{\max}$.
2. If $a < b^{\max}/r^{\max}$, then the RFF is increasing but negatively accelerated.
3. If $a = b^{\max}/r^{\max}$, then the RFF increases linearly from zero to $r^{\max}$.
4. If $a > b^{\max}/r^{\max}$, then the RFF is increasing and positively accelerated.
5. As a approaches infinity, the value of the RFF goes to zero everywhere except at $b = b^{\max}$; it thus approximates a "pulse" on the upper boundary of the interval (0, $b^{\max}$).

If we think of the RFF as describing the "returns" to behavior, then the five cases correspond to (1) returns independent of behavior, (2) marginally decreasing returns, (3) constant returns to behavior, (4) marginally increasing returns, and (5) returns contingent on a specific rate of behavior.

Special cases of the hyperbola have been applied to VI schedules by Baum (1973) and Staddon and Motheral (1978), and the general form was used (also for VI schedules) by Prelec and Herrnstein (1978). Recently, Rachlin (1978) has brought out a different RFF, the power function, which is comparable to the hyperbola in its general properties. In our notation the power RFF would be written as

$$r = db^{a} \tag{6}$$

or, taking reciprocals and logarithms, as

$$\log(1/r) = \log(1/d) + a \log(1/b).$$

A short explication of the difference between the two RFFs is that the hyperbola predicts a linear and the power function a log-linear relationship between interresponse times and interreinforcement times. The magnitude of the discrepancy between these two predictions will depend, of course, on the parameter values and on the observed variation in the response rates. To argue for the superior performance of one or the other function with any particular schedule of reinforcement would digress from the main aims of this article; I shall be content, therefore, to argue mainly for the adequacy of the hyperbola for those schedules on which I will rely for data. Even though the question of empirical superiority is thus set aside, and this question, indeed, may never be satisfactorily resolved because one function may have an advantage with one group of schedules and the other function with another group, it is nevertheless appropriate to point out certain purely formal difficulties that result when the power function is combined with the matching and maximizing response rules. Of these, the most troublesome one is the impossibility of finding an explicit solution (i.e., in elementary functions) for response rates in terms of schedule parameters for even the simple two-schedule feedback system. The interested reader is referred to Appendix A for a complete discussion.

Let us now examine what Assumption 4 implies about (a) real schedules of reinforcement and (b) the nature of reinforcement for unobserved activities.

Real Schedules

My claim is that the hyperbola can adequately account for the relationship between reinforcement and behavior on a variety of schedules of reinforcement.

Time schedules. In pure time schedules the reinforcement is presented independently of behavior. The corresponding RFF is obtained by letting the response requirement in the hyperbolic feedback function equal zero:

$$r = \frac{b}{db + a} = \frac{1}{d} = s. \qquad (7)$$

The rate of reinforcement is constant and equal to s, the reciprocal of the time requirement.

This is a rather trivial RFF; however, it can in principle incorporate the effects of autoshaping, "superstition" experiments, and various classical-conditioning procedures into our feedback model. In addition, it can serve as a rough approximation to the VI schedule in situations where the response rate on the VI is very large compared to the scheduled rate of reinforcement.

Ratio schedules. The RFF for pure ratio schedules is obtained by setting the time requirement, d, equal to zero:

$$r = \frac{b}{db + a} = \frac{1}{a} b = pb. \qquad (8)$$

The received rate of reinforcement is thus strictly proportional to the rate of responding, and the constant of proportionality is the reciprocal of the response requirement a, that is, the probability that any one response will be reinforced.

Interval schedules. The feedback properties of a VI schedule are more difficult to capture mathematically because the obtained rate of reinforcement depends not only on the average rate of responding but also on the microstructure of behavior, for example, on the sequence and distribution of interresponse times. The reason for this is the following: On a VI schedule every interreinforcement interval is composed of two periods. The first is a waiting period during which responding has no effect on reinforcement. After this period is over, reinforcement is set up and only one response is necessary to collect it and complete the cycle. The duration of the waiting period is variable, so it is difficult for the subject to know at any time whether the reinforcement is available or not. In fact, all the VI models considered here share the assumption that the durations of the waiting intervals are sampled from an exponential distribution and in that case the passage of time provides no information at all about reinforcement availability. Because responses emitted before the waiting period has run out have no effect on reinforcement, they are essentially wasted, and different temporal patterns of responding will lead to different fractions of wasted responses and thus to different RFFs. For example, subjects who distribute their response totals evenly over the experimental session will in general obtain higher rates of reinforcement than will subjects who emit the same total number of responses but in a highly inhomogeneous fashion. Hence, any RFF adequate to describe the molar relationship between responding and reinforcement on a VI schedule has to include some additional assumptions about the distribution of interresponse times.

I will now describe three models of VI performance that each induce a particular parametrization of the hyperbolic RFF. The first RFF for VI schedules was proposed by Baum (1973). He reasoned that if all the interresponse times are constant—as in a metronome—then the average interreinforcement interval will be approximately equal to the sum of the average scheduled interreinforcement interval, V, and one half of the average interresponse time:

$$\frac{1}{r} = V + \frac{1}{2} \cdot \frac{1}{b}. \qquad (9)$$

The rate of reinforcement is then a hyperbolic function of the rate of response, with a response requirement of ½ and a time requirement equal to V:

$$r = \frac{b}{Vb + ½}. \qquad (10)$$

The exact RFF for this response pattern is, in fact, somewhat different:

$$r = b(1 - e^{-1/Vb}). \qquad (11)$$

However, when the response rate b is substantially larger than the scheduled reinforcement rate $1/V$, then Equation 10 becomes an excellent approximation to Equation 11.[2]

The second obvious response pattern that one would consider is random responding, in which case the interresponse intervals are independent and exponentially distributed (Staddon & Motheral, 1978). The mean interreinforcement time, then, is simply the sum of the average scheduled reinforcement delay and the average interresponse time:

$$\frac{1}{r} = V + \frac{1}{b},$$

yielding

$$r = \frac{b}{Vb + 1}. \qquad (12)$$

Both of these models are strong idealizations and therefore more reasonable in some situations and less reasonable in others. Their single most serious defect is that they do not allow for the possibility that the subject may be alternating between periods of rapid responding and periods of not responding at all (or perhaps responding on some other schedule). This response pattern, which always characterizes concurrent-schedule performance and very often single-schedule performance as well, creates a bimodal IRT distribution, composed of a small number of long IRTs and a large number of short ones, which is radically different from the two IRT distributions considered previously.

If one did not have an adequate feedback model of this kind of behavior, then Equations 10 and 12 would represent the best models available; however, a third model, developed by Heyman (1977) and Heyman

---

[2] To see that Equation 11 is the exact RFF, recall that in this model all the IRTs have the same duration—$1/b$—and that reinforcement is set up by a Poisson process with an intensity parameter $1/V$. The probability that any one response is reinforced is independent from the reinforcement of any other response, and it is equal to

$$\int_0^{1/b} V^{-1} e^{-t/V} dt = 1 - e^{-1/Vb}.$$

Because the response rate is b, the obtained reinforcement rate will be equal to $b(1 - e^{-1/Vb})$. To see that Equation 10 approximates this function, consider the difference between the scheduled reinforcement rate $1/V$ and the received reinforcement rate r that both functions predict. It follows from Equation 10 that $(1/V) - r = (1/V) - b/(Vb + .5) = 1/[V^2(2b + 1/V)] \cong 1/(2V^2b)$ if $b \gg 1/V$. Equation 11 yields the same quantity as $(1/V) - r = (1/V) - b(1 - e^{-1/Vb})$, and by expanding the exponential as a power series in $(1/Vb)$ one obtains

$$\begin{aligned} &(1/V) - r \\ &\quad = (1/V) - b\left[1 - \sum_{k=0}^{\infty} \left(\frac{-1}{Vb}\right)^k \frac{1}{k!}\right] \\ &\quad = \frac{1}{2!V^2b} - \frac{1}{3!V^3b^2} + \frac{1}{4!V^4b^3} - \cdots \cong \frac{1}{2V^2b}, \end{aligned}$$

if $b \gg 1/V$.

and Luce (1979a), addresses this problem in a relatively sophisticated way and yet can be incorporated into the present theory without increasing its mathematical complexity.

The basic assumption of the Heyman and Luce (1979a) model is that in a concurrent VI-VI situation the subject alternates between periods of responding on one schedule and periods of responding on the other. Because we are now looking at only one schedule in isolation, we may label the periods of responding on it as "on states" and the periods of responding on the other schedule or just resting as "off states." The rate of response during an on state, or the local rate of response, is assumed to be sufficiently high to be treated as continuous; that is, the contribution of the "short" mode of the bimodal IRT distribution to the interreinforcement delay is negligible. Responding is thus completely characterized by the proportion of time spent responding, P.

The durations of individual on states and off states are sampled from two independent exponential distributions with means $1/\mu_1$ and $1/\mu_2$, respectively, and the sampling is unaffected by reinforcement delivery. The proportion of time spent responding is then

$$P = \frac{\mu_2}{\mu_1 + \mu_2}.$$

One can think of μ_2 as the entrance rate to the schedule and μ_1 as the exit rate from the schedule. According to the elegant derivation of Heyman and Luce (1979a), the RFF under such assumptions is

$$r = \frac{P}{V} + \frac{1 - P}{V + I/P}, \tag{13}$$

where V is the average scheduled reinforcement interval and I is equal to $1/(\mu_1 + \mu_2)$. One can rewrite Equation 13 as a hyperbolic function of P:

$$r = \frac{P}{\dfrac{V^2}{I + V}P + \dfrac{IV}{I + V}}. \tag{14}$$

To simplify comparison with other feedback functions, the response measure in this equation can be changed from a dimensionless quantity P to the rate of response measure b by assuming that P corresponds to a fraction, b/k, of the maximum rate of responding, k, that the subject can sustain. Equation 14 then becomes

$$r = \frac{b}{\dfrac{V^2}{I + V}b + \dfrac{kIV}{I + V}} = \frac{b}{db + a}. \tag{15}$$

The time requirement is unchanged by this substitution:

$$d = \frac{V^2}{I + V} \cong V \quad \text{if} \quad V \gg I, \tag{16}$$

whereas the response requirement expressed in the new units becomes

$$a = \frac{kIV}{I + V} \cong kI \quad \text{if} \quad V \gg I. \tag{17}$$

Heyman (1977) has named I the *inertia parameter* because it reflects the subject's tendency to avoid switching from one alternative to another. The utility of this parameter rests on the empirical observation that, unlike other measures of changeover behavior such as frequency of switching, its value for any given subject stays approximately constant over different experimental conditions. Typical values of I run between 1 and 10 sec (Heyman, 1977). In most experiments V will in fact be much larger than I so that the approximate values given by Equations 16 and 17 hold. The response requirement in Equation 17 is the number of responses effectively lost because of sluggish changeover behavior, which corresponds nicely to its mathematical expression as the number of responses that the subject makes in I seconds. For pigeons, k usually lies between 50 and 150 responses per minute. We may conclude, thus, that the response requirement on VI schedules lies in the range of 1 to 30 responses. This estimate of the VI response requirement conflicts sharply with the estimates of the previous two models, which are ½ and 1 response, respectively.

The most useful of the hyperbolic VI feedback functions presented here is the approximation to the Heyman and Luce (1979a) RFF given by Equations 16 and 17:

$$r = \frac{b}{Vb + kI}. \tag{18}$$

Unless it is otherwise specified, this is the one I shall use.

Unobserved Schedules

Assumption 4 states that reinforcement for each unobserved activity is delivered according to a schedule with a hyperbolic RFF. What arguments can be adduced in favor of this assumption?

First, one should realize that the assumption does not limit the number of unobserved activities; all it entails is that there exist one, not necessarily unique, way of disaggregating background activity so that each individual activity is reinforced by a schedule with a hyperbolic RFF.

Second, there are several classes of molecular models of background or spontaneous reinforcement that yield, on the aggregate level, the hyperbolic RFF. Real schedules of reinforcement provide one class of such models. If the unobserved activity—say scratching—is composed of discrete individual actions—scratches—that have some probability of being spontaneously reinforced, then the schedule is a VR; if each individual successful scratch is followed by a refractory period of variable duration, then the schedule is a VI, and so on.

A different class of molecular models can be constructed along the following lines: Assume that the value of reinforcement for each individual action is itself variable, specifically, that it is an increasing function of the time since the previous reinforced action of that same class (e.g., scratching). The reinforcement of one response initiates a temporary period of satiation, which then gradually wears off. To describe this process mathematically, let $\zeta(t)$ be the magnitude of satiation or habituation after t seconds and $\rho^{\max}$ be the value of reinforcement for one response when there is no satiation. The value of reinforcement after t seconds have elapsed without a response may then be defined as

$$\rho(t) = \rho^{\max}\{1 - [\zeta(t)/\zeta(0)]\} \quad (d\zeta/dt \leq 0). \tag{19}$$

Immediately after reinforcement, satiation is complete and responding is not reinforced; as $\zeta(t)$ decreases to zero with the passage of time, $\rho(t)$ will approach its maximum value $\rho^{\max}$.

Here, as in the analysis of schedules of reinforcement, I wish to derive the relationship between the average rate of response and the average rate of reinforcement (weighted by the value of individual reinforcing events). In order to do this, the form of the function $\zeta(t)$ and the molecular properties of responding need to be specified. Assume that satiation decays exponentially, that is, $\zeta(t) = e^{-t/\tau}$, and that responding is a random or Poisson process. It can then be shown that weighted reinforcement rate is a hyperbolic function of response rate:[3]

$$r = \frac{(\rho^{\max})b}{\tau b + 1}. \tag{20}$$

The maximum rate of reinforcement—$\rho^{\max}/\tau$—is proportional to the magnitude of reinforcement when there is no satiation ($\rho^{\max}$), divided by a quantity proportional to the half-life of the satiation decay process (τ).

This model illustrates a more general point. When satiation dissipates rapidly ($\tau \rightarrow 0$), the RFF for this model becomes linear; the equivalent schedule, therefore, is a VR. When satiation is persistent ($\tau \rightarrow \infty$), the RFF becomes flat; the equivalent schedule is now closer to a VI. One can speculate that strong satiation is more likely to occur if the animal is not deprived of the particular activity and, conversely, is less likely to occur if the animal is deprived. Given that most of the background activities in the experimental chamber (walking about, grooming, etc.) are also possible outside of

---

[3] In order to derive Equation 20, first define two random variables, IRT (the distribution of interresponse times) and Q (the distribution of reinforcement magnitudes). My assumptions are

(a) $\text{Prob}\{\text{IRT} \geq x\} = 1 - e^{-bx}$

(b) $Q = \rho^{\max}(1 - e^{-\text{IRT}/\tau})$.

I define the weighted rate of reinforcement as equal to the expected magnitude of reinforcement divided by the expected duration of the interresponse time:

$$r = [E(\text{IRT})]^{-1}E(Q) = bE(Q)$$
$$= b\int_0^\infty \rho^{\max}(1 - e^{-x/\tau})(be^{-bx})dx = \frac{\rho^{\max}b}{\tau b + 1}.$$

it and that, consequently, the animal is not deprived of them, one may expect background schedules to be better represented by VI rather than VR schedules. Background activity, in other words, should exhibit sharply decreasing marginal returns (Staddon, 1980).

The last argument for the hyperbolic representation of background schedules exploits a convenient mathematical property of the hyperbolic function, namely, that it is closed under composition. This means that schedules which simultaneously control the value and the probability of reinforcement can still have hyperbolic RFFs.

Let us temporarily draw a distinction between the frequency and the rate of reinforcement that such a schedule would provide. Frequency (x) denotes the number of reinforcing events—regardless of their value—per unit time; rate (r)—as in the previous model of satiation—denotes the frequency of reinforcement multiplied by the average value of reinforcement per reinforcing event. The probabilistic component of the schedule will then impose a hyperbolic relation between x and b, $x = b/(db + a)$, whereas the component that models the local satiation process will impose a hyperbolic relation between r and x, for example, $r = \rho^{\max}x/(x\tau + 1)$. The RFF is the composition of these two functions, and it happens to be, again, hyperbolic:

$$r = \frac{\rho^{\max}b}{(\tau + d)b + a}.$$

The construction of additional models that satisfy the hyperbolic RFF would be a misdirected exercise. We simply do not know enough about background contingencies to build persuasive models. This last property suggests, however, that future models will be able to expand in two directions simultaneously—incorporating probabilistic and nonprobabilistic variation—and still stay within the limits imposed by the hyperbolic RFF.

The Matching and Maximizing Response Rules

To complete the theory only the response rule that the subject is following need be specified. Assumptions 5(M) and 5(O) define the two possibilities, matching and maximizing.

ASSUMPTION 5(M). *At equilibrium, the response share of each activity, b_i/k_i, equals the weighted relative rate of reinforcement for that activity*:

$$b_i/k_i = \frac{\rho_i r_i}{\sum_{j=1}^{n} \rho_j r_j}, \quad i = 1, \ldots, n.$$

ASSUMPTION 5(O). *At equilibrium, the total weighted rate of reinforcement,*

$$\sum_{j=1}^{n} \rho_j r_j,$$

is maximized.

The maximization assumption can also be formulated as a set of simultaneous equations. Using ordinary techniques of constrained maximization (Intriligator, 1971), we form the Lagrangean function,

$$\mathrm{L}(b_1, \ldots, b_n, \lambda) = \sum_{j=1}^{n} \rho_j r_j(b_j) - \lambda[\sum_{j=1}^{n} (b_j/k_j) - 1],$$

and differentiate it with respect to $b_1, \ldots, b_n$ and λ. This produces n equations of the form $\rho_j dr_j/db_j = \lambda/k_j$. Adding all n equations shows that the constant λ is equal to

$$1/n(\sum_{j=1}^{n} k_j\rho_j dr_j/db_j).$$

The necessary conditions for a maximizing response distribution can then be written as

$$k_i\rho_i \frac{dr_i}{db_i} = \frac{1}{n}(\sum_{j=1}^{n} k_j\rho_j dr_j/db_j), \quad i = 1, \ldots, n. \tag{21}$$

Earlier, I said that the constants ρ_i should be interpreted as marginal utilities. The reason for such an identification is this: Were the subject maximizing some general utility function $\mathrm{U}(r_1, \ldots, r_n)$ defined for combinations of reinforcement rates, his or her behavior would conform to an alternative set of equations that could be obtained from the original set (Eq. 21) by systematically replacing reinforcement values ρ_i with marginal utilities $\partial\mathrm{U}/\partial r_i$:

$$k_i \frac{\partial U}{\partial r_i}\frac{dr_i}{db_i} = \frac{1}{n}\left[\sum_{j=1}^{n} k_j \partial U/\partial r_j(dr_j/db_j)\right],$$

$$i = 1, \ldots, n.$$

The linear utility function

$$U = \sum_{j=1}^{n} \rho_j r_j ,$$

which is implied by Assumption 5(O), is a first-order approximation to the general utility function $U(r_1, \ldots, r_n)$, and like all approximations, it fares better in some contexts than in others. In making this assumption I do not intend to deny the fact that in many experiments the marginal utilities of different reinforcers interact—that a high rate of one reinforcer will inflate or depress the marginal utility of another—but simply to set up an ideal standard by which the relevance of individual experimental findings can be estimated. The linear utility function should be a reasonable model when either (a) all schedules deliver the same reinforcer or (b) the experimental session is very brief. In practice, the second condition is hard to satisfy, so I shall fall back on the first one and only consider experiments in which all real schedules deliver a physically identical reinforcer—invariably, small bits of food. It is true that the marginal utility of background reinforcers may still vary, but no harm comes from that as long as the relationship between rate and marginal utility resembles a hyperbola. Besides, in most experiments with food reinforcement, the animal's extreme hunger diminishes the overall importance of background activities and makes precise modeling of background reinforcement unnecessary.

The assumption of constant reinforcement values has the same significance for matching theory as it does for maximization theory when allowances are made for a different and perhaps less established vocabulary. The role of the marginal utility functions $\partial U/\partial r_i(r_1, \ldots, r_n)$ is taken over by the reinforcement value functions $\rho_i(r_1, \ldots, r_n)$. The only difference between the two sets of functions is that the marginal utility functions must be integrable (there must exist a utility function U that when differentiated, yields the marginal utility functions) whereas the reinforcement value functions need not be. A possibly important consequence of this distinction is that the marginal utility functions create a symmetric substitutability/complementarity dimension; that is, reinforcer A is a substitute for reinforcer B only to the extent that B is a substitute for A, whereas the reinforcement value functions can accomodate the odd pair A, B such that A is a substitute for B but B is not a substitute for A.

Before continuing it will prove expedient to suppress the constants k_i and ρ_i, whose only purpose was to show how differences among response topographies and among reinforcing events could be dealt with. Because these constants do not interact in any interesting way with the main qualitative implications of the theories, they are from this point on superfluous.

Conversion constants can be eliminated by rescaling all parameters and variables with respect to one observed pair of response and reinforcement measures (Herrnstein, 1974). The correct set of rescaling transformations, given that the first activity is the reference, replaces b_i, r_i, a_i, and d_i, with $(k_i/k_1)b_i$, $(\rho_1/\rho_i)r_i$, $(\rho_i/\rho_1)(k_i/k_1)a_i$, and $(\rho_i/\rho_1)d_i$. By this substitution, all reinforcement value constants ρ_i drop out, and so do all but one of the maximum response rate constants k_i. Because the choice of the first schedule for the referent schedule was arbitrary, I will write this remaining constant (k_1) simply as k.

Is is now possible to formulate alternative definitions of matching and maximizing behavior. These new definitions, although following strictly from the more basic ones given by Assumption 5(M) and Assumption 5(O), show what the conflict between the matching and maximizing response rules reduces to after all the problems of response and reinforcement scaling have been set aside. From Assumption 5(M) and Equation 21 we may immediately derive the following assumptions.

ASSUMPTION 5(M)′. *After rescaling, the average (or local) rates of reinforcement are the same on every schedule*:

$$\frac{r_i}{b_i} = \lambda^{(m)}, \quad i = 1, \ldots, n,$$

where

$$\lambda^{(m)} = \sum_{j=1}^{n} r_j/k.$$

ASSUMPTION 5(O)′. *After rescaling, the marginal rates of reinforcement are the same on every schedule*:

$$\frac{dr_i}{db_i} = \lambda^{(o)}, \quad i = 1, \ldots, n,$$

where

$$\lambda^{(o)} = \sum_{j=1}^{n} (dr_j/db_j)/n.$$

So, we have returned to the basic question posed at the end of the first section: What are the controlling variables? Are they the average (local, if b_i measures time) or the marginal rates of reinforcement? To pose the question this way is not to commit oneself to a mechanistic conception of controlling variables, or for that matter to any particular underlying process by which one or the other response rule could be realized; it is only to make use of the fact that the equilibrating behavior of a subject who matches or maximizes must, in the long run, look as if it is being controlled by the appropriate set of variables and nothing else.

Basic Properties of Matching and Maximizing Behavior

Assumptions 1–5 constrain responding in several important ways. In this section I shall examine the general properties of matching and maximizing performance; the analysis of specific experiments is postponed until the next section.

Pairwise Properties

The first three theorems connect the present models with some more traditional choice-theoretic issues. Let us start with some definitions. First, behavior on two fixed schedules i and j is defined as *pairwise linear* if there exist constants ω_{ij} and χ_{ij} such that

$$b_i = \omega_{ij} b_j + \chi_{ij}. \tag{22}$$

Second, behavior on two fixed schedules i and j obeys the *constant-ratio rule* if it is pairwise linear and if the constant χ_{ij} is equal to zero:

$$b_i = \omega_{ij} b_j. \tag{23}$$

This property is a translation of Luce's choice axiom (Luce, 1959) into the response-rate idiom. It states that the relative rate of response (b_i/b_j) for any two alternatives i and j is a constant, independent of the structure and availability of other alternatives. Changes in context—that is, the addition of some schedules or the deletion of others—could, in this case, change the absolute rates for response, b_i and b_j, but the ratio, b_i/b_j, would be preserved. Pairwise linearity can then be taken as a possible first line of retreat, so to speak, from the constraints of the choice axiom. The relationship between b_i and b_j is still sharply restricted, but the ratio, b_i/b_j, can now take on practically any value.

Third, let us say that behavior on a fixed schedule i is independent of behavior on a fixed schedule j if pairwise linearity holds and the constant ω_{ij} is equal to zero:

$$b_i = \chi_{ij}. \tag{24}$$

The implications of the matching and maximizing response rules for pairwise properties can be assessed directly from Assumptions 5(M)′ and 5(O)′. Matching implies that the local rates of reinforcement (r/b) on any two schedules are equal, given that the RFFs for the two schedules are hyperbolic, that is,

$$\frac{1}{d_i b_i + a_i} = \frac{1}{d_j b_j + a_j}.$$

Maximization implies that the marginal rates are equal, that is,

$$\frac{a_i}{(d_i b_i + a_i)^2} = \frac{a_j}{(d_j b_j + a_j)^2}.$$

Solving for b_i in these two equations then proves the first theorem.

THEOREM 1. *If the subject matches or maximizes, then pairwise linearity holds for all pairs of schedules. Matching implies*

$$b_i = (d_j/d_i) b_j + (a_j - a_i)/d_i.$$

Maximizing implies

$$b_i = (a_i^{1/2} d_j / a_j^{1/2} d_i) b_j + a_i^{1/2}(a_j^{1/2} - a_i^{1/2})/d_i.$$

This theorem has a special virtue in that it can be tested even if data on reinforcement rates are not available. Consider two activities such as wheel running and resting for the rat. They could be treated as two partially observable schedules—schedules that have associated response measures but lack reinforcement measures. According to Theorem 1, if we measure the amount of time spent resting and running while simultaneously varying the parameters controlling a third schedule—say a schedule of food reinforcement—then the amount of time spent running should be a linear function of the amount of time spent resting.

As a byproduct of Theorem 1, one can derive a simple preliminary characterization of the relationship between matching and maximizing behavior. By substituting for d_i and d_j in terms of r_i and r_j in the maximization part of Theorem 1, one obtains

$$b_i = (a_i^{1/2} r_i / a_j^{1/2} r_j) b_j \,.$$

Maximizing, in other words, is a form of biased matching (Baum, 1974; Rachlin, 1978): $(b_i/b_j) = (a_i/a_j)^{1/2}(r_i/r_j)$. Generalizing this over all n schedules, we have[4]

THEOREM 2. *The maximizing response rule is equivalent to a weighted matching response rule*:

$$\frac{b_i^{(o)}}{\sum_{j=1}^{n} b_j^{(o)}} = \frac{a_i^{1/2} r_i^{(o)}}{\sum_{j=1}^{n} a_j^{1/2} r_j^{(o)}} \,.$$

This is matching behavior, but with respect to a systematically modified set of reinforcement rates. The subject responds more on those schedules that have a higher associated response requirement. Therefore, the extent to which matching and maximizing behavior will differ depends exclusively on the distribution of the response-requirement values among the various schedules in the environment. When the response requirements are roughly equal, then matching and maximizing lead to similar predictions.

Let us now apply Theorem 1 to particular pairs of schedules. If Schedule 1 in the Theorem is a VI, with time and response requirements of V and kI, and Schedule 2 is a VR, with time and response requirements of 0 and a_2, then the matching and maximizing rates of response on the VI, $b_1^{(m)}$ and $b_1^{(o)}$, are

$$\begin{aligned} b_1^{(m)} &= (a_2 - kI)/V, \\ b_1^{(o)} &= [(a_2 kI)^{1/2} - kI]/V. \end{aligned} \tag{25}$$

The absence of $b_2^{(m)}$ and $b_2^{(o)}$ in these two equations indicates that under both response rules, response rate on the VI schedule is not functionally constrained by the response rate on the VR; in other words, behavior on the VI is independent of behavior on the VR. Because the derivation of this result from Theorem 1 only required that d_2 be equal to zero (e.g., that Schedule 2 is a ratio schedule), one can state, more generally, that behavior on any schedule is not functionally constrained by behavior on a VR (it is functionally constrained, however, by the ratio schedule's response requirement, denoted here by a_2).

With either the matching or the maximizing response rule, the ratio schedule is a stabilizing schedule in the following sense: In an environment not containing a ratio schedule, behavior on any single schedule will be sensitive to variation in the *choice set*—the set of concurrently available schedules; however, when a ratio schedule is added to the experiment, then any additional changes in the composition of the choice set will only affect responding on that ratio schedule. For that same reason one should expect ratio-controlled responding to be more variable; it has to absorb entirely the impact of fluctuations in the deprivation level (which are represented in the theory as fluctuations in the parameters of background schedule RFFs).

The second example that I shall consider is the case of a VI–VI combination. If the mean scheduled delays on the two schedules are denoted by V_1 and V_2, and their common response requirement by kI, then Theorem 1 implies

---

[4] The superscripts (m) and (o) will be employed with results based on matching (Assumptions 1-5[M]) and maximizing (Assumption 5[O]), respectively. They will be omitted if the response rule is clearly identified by the context.

$$
\begin{aligned}
b_1^{(\mathrm{m})} &= (V_2/V_1)b_2^{(\mathrm{m})}, \\
b_1^{(\mathrm{o})} &= (V_2/V_1)b_2^{(\mathrm{o})}.
\end{aligned}
\tag{26}
$$

We notice two things about these equations: First, they are identical; second, the constant-ratio rule holds. The VI pair thus illustrates the point of the next theorem: When matching and maximizing equilibria coincide, then the constant-ratio rule holds, and conversely, when either matching or maximizing implies the constant-ratio rule, then two sets of equilibria are identical.

THEOREM 3. *Given Assumptions 1–4, any two of the following three conditions imply the remaining one: (1) The subject matches. (2) The subject maximizes. (3) The subject obeys the constant-ratio rule for all pairs of schedules.*

Assumption 4 is not a necessary precondition for this theorem; it is sufficient in its stead to assume that each RFF (a) is continuous and twice differentiable, (b) is increasing but marginally decreasing, and (c) gives no reinforcement in the absence of responding. Taken together, these three fairly innocuous restrictions are strong enough to allow a proof of Theorem 3 (see Appendix B).

For hyperbolas, of course, this theorem hardly requires proof. By Theorem 2 we know that the matching and maximizing response distributions are identical if and only if all the response requirements are identical; by Theorem 1 we see that the constant-ratio rule has the same prerequisite. Identical response requirements cancel out, and performance is determined solely by the time requirements:

$$
b_i^{(\mathrm{o})} = b_i^{(\mathrm{m})} = \frac{{d_i}^{-1_k}}{\sum_{j=1}^{n} {d_j}^{-1}}. \tag{27}
$$

This theory would be very different, indeed, were it based solely on schedules with identical response requirements. First, it would be isomorphic to the original choice model (Luce, 1959), and consequently, it would propose and defend relative response rate as the stable measure of preference between two schedules or, more generally, between two alternatives. Second, it would be insensitive to the tension between matching and maximizing response rules; depending on the context, one or the other could be used as interchangeable normative definitions of rational behavior. Third, it would find a concept like the RFF superfluous. Each schedule or activity could be assigned a number—a value—proportional to d^{-1} and, as the relative value of the schedule would completely determine its response share, the RFF would be replaced by the theoretically simpler but equally predictive notion of value.

The existence of a ratio scale—along which alternatives may be ranked in such a way that, barring exclusive preference, the probability (relative frequency) of choosing any one out of a set of them is equal to the value of that alternative divided by the sum of the values of the alternatives in the set (viz., Eq. 27)—is one of the more significant implications of Luce's choice axiom. In the past the relationship between matching and the choice axiom has often been misunderstood, and the dissatisfaction some authors have felt about the matching law was no doubt based in part on the belief that matching would in some way force upon them the various theoretical austerities of the choice axiom. The most recent example of this can be found in the article by Rachlin et al. (1980) where they state that "Luce's (1959) choice axiom . . . is equivalent to a simple form of Herrnstein's (1961) matching law" (p. 355) from which they then set up, for comparison with maximization theory, an unnecessarily restricted version of the matching rule.

It may be useful to try to reconstruct how this identification of matching with the choice axiom could have come about. The trouble began, I suppose, when the matching law came to be written as a relation between the times allocated to different alternatives (T_i) and the obtained values of reinforcement obtained from these alternatives (V_i):

$$
\frac{\mathrm{T}_i}{\sum_{j=1}^{n} \mathrm{T}_j} = \frac{V_i}{\sum_{j=1}^{n} V_j}.
$$

Having defined matching in this way, it becomes tempting to draw the inference that the ratios T_i/T_j, being equal to V_i/V_j, will stay constant regardless of context and so assimilate matching to the choice axiom. To see that this is a false inference, one has to remember what the values V_i stand for. They are not v-scale values in the choice axiom sense; they are, instead, functions of the physical parameters of reinforcement, such as rate, amount, immediacy, and so on (Baum & Rachlin, 1969). At least some of the physical parameters, most certainly rate or overall duration of reinforcement, will depend on time invested and so will place the obtained values into a functional dependence on behavior. But, if the V_i are then written as they should be—as $V_i(T_i)$ or, better yet, as $V_i(T_1, \ldots, T_n)$—then there are no longer any grounds for asserting that the ratios in question, T_i/T_j, will remain constant as the set of alternatives changes (Prelec & Herrnstein, 1978). It is only through a careful examination of the feedback relations (here $V_i[T_i]$) that this claim can be properly evaluated.

With certain classes of schedules, such as the variable-interval class, the constant ratio rule will hold, and performance on these schedules can be predicted by a straightforward application of the scaling techniques associated with the choice axiom. But even in circumstances so favorable to an identification of matching and the choice axiom, it is self-contradictory to interpret the values that this scaling exercise would assign to individual schedules as the values of obtained reinforcements. For the V_i's to stand for obtained reinforcements, some connection, however remote and hypothetical, must exist between them and the physical parameters of reinforcement, and the scale values obtained by an application of the choice axiom do not, almost by definition, describe any aspect of physical reality. The d_i^{-1} in Equation 27 will accurately describe performance, given that all response requirements are equal, but being constants, they will not stand in any determinate relation to the sole physical parameters—the obtained rates of reinforcement r_i—which can go up and down depending on the competition provided by concurrently available alternatives.

The main implication of Theorem 3 for all this is that the matching and maximizing response rules are similarly related to the constraints of the choice axiom: They both support them in certain classes of choice situations and contradict them in all others. Appending a value structure to the two response rules will not change the substance of this result; it will only mean that the theorem will have to be applied to functions that are in fact compositions of two functions: one that shows how the physical parameters of reinforcement depend on responding and another that shows how value or utility in turn depends on these physical quantities. A violation of the choice axiom can be used to support one theory against another only to the extent that one has already built into them some asymmetry in the functional relation of the relevant value variables to responding. This, in effect, is what Rachlin et al. (1980) have done by not letting the obtained values V_i depend on responding in the matching model but allowing for these same dependencies in the maximizing model with which the matching one was being compared.

Aggregation Theorems

The next theorem shows that several schedules can be treated as one composite schedule.

THEOREM 4(M). *If the subject matches, then the form of the RFF is preserved under aggregation of response and reinforcement classes. For any set of schedules Q, let*

$$b_Q = \sum_{i \in Q} b_i \quad \textit{and} \quad r_Q = \sum_{i \in Q} r_i\,;$$

then

$$r_Q = \frac{b_Q}{d_Q^{(m)} b_Q + a_Q^{(m)}},$$

where

$$d_Q^{(m)} = \Big(\sum_{i \in Q} d_i^{-1}\Big)^{-1}$$

and

$$a_Q^{(m)} = \Big(\sum_{i \in Q} a_i d_i^{-1}\Big)\Big(\sum_{i \in Q} d_i^{-1}\Big)^{-1}.$$

It would be difficult to overestimate the usefulness of this property. It allows us to aggregate response classes in any way without disturbing the hyperbolic form of the RFF. The RFF for the larger response classes will always be hyperbolic, provided that the RFFs at the most disaggregated level are hyperbolic.

First, Theorem 4(M) can be applied to some specific cases. Imagine that we want to combine responding maintained by a VI schedule (b_1) and responding maintained by a VR schedule (b_2) into one response category ($b_{1+2} = b_1 + b_2$). What would the resulting RFF be? Applying the theorem one can derive the aggregate time and response requirements (d_{1+2} and a_{1+2}) from the individual ones (which are $d_1 = V$, $a_1 = kI$, $d_2 = 0$, and $a_2 = a_2$):

$$d_{1+2} = \lim_{d_2 \to 0} (V^{-1} + d_2^{-1})^{-1} = 0 \quad (28)$$

$$a_{1+2} = \lim_{d_2 \to 0} (kIV^{-1} + a_2 d_2^{-1})/(V^{-1} + d_2^{-1})$$

$$= a_2 .$$

The aggregate RFF is identical to the RFF of the ratio schedule: $r_{1+2} = b_{1+2}/a_2$. Because the derivation of this fact did not depend on the values of the VI schedule parameters, we may, by induction, generalize it in the following way: If the aggregate schedule contains a ratio schedule, and if the subject matches, then the aggregate schedule will also be a ratio schedule.

Second, the aggregation process can be carried to the limit and aggregate all responding into one class. The associated hyperbolic RFF shows how the total rate of reinforcement received from all the schedules in the environment should covary with k—the amount of behavior available to the organism. Because k is fixed, this function is not an RFF in the strict sense of the term. It does, however, compactly characterize the essential properties of the complete feedback system. Applying the aggregation rule to all n schedules, we can derive the total time requirement a_T and the total response requirement d_T:

$$d_T = (\sum_{i=1}^{n} d_i^{-1})^{-1} \text{ and } a_T = (\sum_{i=1}^{n} a_i d_i^{-1})/d_T .$$

The RFF for the total amount of reinforcement is then

$$r_T = \frac{k}{d_T k + a_T} . \quad (29)$$

How should the parameters a_T and d_T be interpreted? The total time requirement is the harmonic mean of the individual time requirements. Its reciprocal (s_T) is the maximum rate of reinforcement that the environment could provide if responding was unconstrained:

$$s_T = d_T^{-1} = \lim_{k \to \infty} r_T(k).$$

The total response requirement a_T is the sum of the response requirements for all the individual schedules, weighted by the s_i, which gives, loosely speaking, the relative importance of Schedule i:

$$a_T = \frac{\sum_{i=1}^{n} s_i a_i}{\sum_{i=1}^{n} s_i} .$$

This aggregate parameter describes the extent to which reinforcement is ratio-coupled to behavior. When the number is small, then the environment delivers reinforcement largely independent of the behavior; when it is large, then the animal must work for its reinforcers.

Maximization does not ensure that the form of the hyperbolic RFF is preserved under aggregation of schedules; rather, it ensures the preservation of the form of the marginal RFF. This function, which, if the RFF is hyperbolic, has the form

$$\frac{dr}{db} = \frac{a}{(db + a)^2} , \quad (30)$$

shows how the marginal rate of reinforcement depends on the rate of response.

THEOREM 4(O). *If the subject maximizes, then the form of the marginal RFF is preserved under aggregation of response and reinforcement classes. For any set of schedules Q, let*

$$b_Q = \sum_{i \in Q} b_i \quad \textit{and} \quad r_Q = \sum_{i \in Q} r_i ;$$

then

$$\frac{dr_Q}{db_Q} = \frac{a_Q}{(d_Q^{(o)} b_Q + a_Q^{(o)})^2} ,$$

where

$$d_Q^{(o)} = (\sum_{i \in Q} a_i d_i^{-1})(\sum_{i \in Q} a_i^{1/2} d_i^{-1})^{-2}$$

and

$$a_Q^{(o)} = (\sum_{i\in Q} a_i d_i^{-1})^2 (\sum_{i\in Q} a_i^{1/2} d_i^{-1})^{-2}.$$

This is strictly a weaker property than the one derived in Theorem 4(M).

Response and Demand Functions

A complete solution to the system of equations defined by Assumptions 1-5 is a set of response functions $\{f_i\}$, which show how response rates vary with the parameters of the system:[5]

$$b_i = f_i(a_j;\, d_j;\, k).$$

Alternatively, one may characterize the solution by means of a set of demand functions $\{g_i\}$, which show how the received rates of reinforcement vary with these same parameters:

$$r_i = g_i(a_j;\, d_j;\, k).$$

The response functions can be easily derived from the demand functions, and vice versa, because they are connected with a hyperbolic RFF:

$$g_i = \frac{f_i}{d_i f_i + a_i}.$$

The label *demand function* is a concession to economics; it is appropriate in this context because the functions g_i indicate how much of a certain reinforcer will be "demanded" by the subject, given the "costs" as summarized by the vector of parameter values. In economics an analogous quantity—the quantity of goods purchased at a given price—is the primary variable of interest. The psychologist, however, is more concerned with the response distribution, so it is more convenient to work with the response functions.

THEOREM 5(M). *The matching response and demand functions are*

$$b_i^{(m)} = \frac{k + \sum_{j=1}^{n} (a_j - a_i) d_j^{-1}}{d_i \sum_{j=1}^{n} d_j^{-1}}$$

and

$$r_i^{(m)} = \frac{k + \sum_{j=1}^{n} (a_j - a_i) d_j^{-1}}{d_i (k + \sum_{j=1}^{n} a_j d_j^{-1})}.$$

THEOREM 5(O). *The maximizing response and demand functions are*

$$b_i^{(o)} = \frac{k + \sum_{j=1}^{n} (a_j^{1/2} - a_i^{1/2}) a_j^{1/2} d_j^{-1}}{a_i^{-1/2} d_i \sum_{j=1}^{n} a_j^{1/2} d_j^{-1}}$$

and

$$r_i^{(o)} = \frac{k + \sum_{j=1}^{n} (a_j^{1/2} - a_i^{1/2}) a_j^{1/2} d_j^{-1}}{d_i (k + \sum_{j=1}^{n} a_j d_j^{-1})}.$$

The cumbersome mathematical form of these functions conceals an important underlying relationship: There exists a mapping of the parameter space that transforms the matching response function into the maximizing response function. In other words, matching behavior in one experiment is simultaneously maximizing behavior for another, different experiment:

$$f_i^{(m)}(a_j;\, d_j;\, k) = f_i^{(o)}(a_j^2;\, a_j d_j;\, k) \qquad i = 1, \ldots, n \tag{31}$$

or

$$f_i^{(o)}(a_j;\, d_j;\, k) = f_i^{(m)}(a_j^{1/2};\, a_j^{-1/2} d_j;\, k) \qquad i = 1, \ldots, n. \tag{32}$$

The maximizing response function has the same form as the matching function except that the parameter values have been systematically altered: The time requirement d_j in $f_i^{(m)}$ appears as $a_j^{-1/2} d_j$ in $f_i^{(o)}$, whereas the response requirement a_j appears as $a_j^{1/2}$. Hence, all observations made about the matching response functions also follow about the maximizing response function, with the appropriate transformation of parameters.

The response function is a useful tool only when many parameters vary simultaneously. When this is not the case—when some schedules have parameters that are fixed throughout the observation period—then there frequently exist simpler descriptions of matching and maximizing behavior.

---

[5] The notation $f_i(a_j;\, d_j;\, k)$ is a shorthand for $f_i(a_1, \ldots, a_n, d_1, \ldots, d_n, k)$.

The response and demand functions simplify considerably if all schedules but one are fixed, that is, if only one pair of a_i and d_i can vary. By the aggregation theorem we can treat all the schedules other than schedule i as a single schedule. This aggregate background schedule will be called the *complementary schedule* of Schedule i and will be labeled by the subscript i'. The parameters $a_{i'}$ and $d_{i'}$ are, respectively, the aggregate response and time requirements of the complementary schedule. $d_{i'}^{-1} = s_{i'}$ is also the maximum rate of reinforcement that schedules other than Schedule i can deliver.

For matching, two new parameters,

$$\alpha^{(\mathrm{m})} = k + \sum_{j \neq i} a_j d_j^{-1} = k + a_{i'} d_{i'}^{-1} \tag{33}$$

and

$$\beta^{(\mathrm{m})} = \sum_{j \neq i} d_j^{-1} = d_{i'}^{-1},$$

allow us to rewrite Theorem 5(M) as

$$b_i = \frac{\alpha^{(\mathrm{m})} - \beta^{(\mathrm{m})} a_i}{1 + \beta^{(\mathrm{m})} d_i} \tag{34}$$

$$r_i = \frac{\alpha^{(\mathrm{m})} - \beta^{(\mathrm{m})} a_i}{a_i + \alpha^{(\mathrm{m})} d_i}. \tag{35}$$

For maximizing, the substitution of

$$\alpha^{(\mathrm{o})} = k + \sum_{j \neq i} a_j d_j^{-1} = k + a_{i'}^{(\mathrm{o})} (d_{i'}^{(\mathrm{o})})^{-1} \tag{36}$$

and

$$\beta^{(\mathrm{o})} = \sum_{j \neq i} a_j^{1/2} d_j^{-1} = (a_{i'}^{(\mathrm{o})})^{1/2} (d_{i'}^{(\mathrm{o})})^{-1}$$

into Theorem 5(O) leads to

$$b_i = \frac{\alpha^{(\mathrm{o})} - \beta^{(\mathrm{o})} a_i^{1/2}}{1 + \beta^{(\mathrm{o})} d_i a_i^{-1/2}} \tag{37}$$

and

$$r_i = \frac{\alpha^{(\mathrm{o})} - \beta^{(\mathrm{o})} a_i^{1/2}}{\alpha^{(\mathrm{o})} d_i + a_i}. \tag{38}$$

These equations are the simplest way to present predictions of performance on single-schedule experiments, and these predictions will be compared to data later on. However, if the subject matches, then a yet simpler statement about performance on these schedules can be made.

THEOREM 6. *If the subject matches, and if all schedules except Schedule i are fixed, then the following relation holds*:

$$b_i = \frac{\alpha^{(\mathrm{m})} r_i}{r_i + \beta^{(\mathrm{m})}}.$$

This theorem is really a response rule: It states how the environment, as parametrized by the rate of reinforcement, controls behavior. The curious thing about it is that it has the same mathematical form as the RFF, which defines how behavior controls the environment. Insofar as one schedule is concerned, equilibrium behavior is defined by the intersection of two hyperbolas, those of the RFF and Theorem 6.

These applications of the aggregation theorem to single-schedule performance make the problem of background reinforcement theoretically and practically more manageable. That the experimenter does not know the number of background schedules or what their particular properties are is not troublesome as long as the assumption that each individual schedule has a hyperbolic RFF seems plausible to him or her. The estimation of parameters for background schedules is now simpler too: It is not necessary to entertain separately the hypotheses that there are one, two, three, or more background schedules; it is sufficient to estimate the two aggregate parameters $a_{i'}$ and $d_{i'}$. Whether background reinforcement is controlled by one hyperbolic RFF or one thousand should make no difference whatsoever.

Long-Run Determinants of Extinction

The theorems discussed so far have all depended critically on the assumption that response rates are greater than zero and less than k (Assumption 2). If any response rate drops to zero or hits k, then the theorems no longer hold.

This assumption is not a problem unless there are some extremely strong or extremely weak schedules in the system. It is, in fact, possible for a schedule to be so poor that it cannot simultaneously maintain any responding (Assumption 2) and satisfy the matching or maximizing conditions (Assumption 5); similarly, a schedule can be so generous that it will either capture all responding or violate the relevant version of

Assumption 5. The five original assumptions of the theory are thus incompatible with certain regions in the parameter space; the parametrizations of the individual schedules cannot, in some sense, be too different in magnitude from one another. The next theorem defines this restriction precisely.

THEOREM 7(M). *For each Schedule i, Assumptions 1–5(M) imply*

$$|2(a_i - a_{i'}^{(\mathrm{m})}) + k(d_i - d_{i'}^{(\mathrm{m})})| < k(d_i + d_{i'}^{(\mathrm{m})}).$$

THEOREM 7(O). *For each Schedule i, Assumptions 1–5(O) imply*

$$\begin{aligned}|2[a_i^{1/2} - (a_{i'}^{(\mathrm{o})})^{1/2}] \\ + k[d_i a_i^{-1/2} - d_{i'}^{(\mathrm{o})}(a_{i'}^{(\mathrm{o})})^{-1/2}]| \\ < k[d_i a_i^{-1/2} + d_{i'}^{(\mathrm{o})}(a_{i'}^{(\mathrm{o})})^{-1/2}].\end{aligned}$$

To test if all response rates will be positive, therefore, the parameters of every schedule i have to be compared against the parameters of its complementary schedule i'.

For matching, the nature of the comparison is simple: Theorem 7(M) can be rewritten as a pair of inequalities:

$$kd_i < a_i - a_{i'}^{(\mathrm{m})} < kd_{i'}^{(\mathrm{m})}.$$

If schedule i has the greater response requirement, then the difference between the two response requirements must not exceed $kd_{i'}$ responses or responding on the schedule will cease; if, on the other hand, the complementary schedule has the greater response requirement, then the difference between the two must be less than kd_i responses or responding on the complementary schedule will cease (see Appendix C).

The same basic principle holds for maximizing, after the usual transformation of the parameter space:

$$a_i \rightarrow a_i^{1/2} \quad \text{and} \quad d_i \rightarrow a_i^{-1/2} d_i .$$

Empirical Findings

Single-Schedule Data

The two models presented here make a number of straightforward and sometimes conflicting predictions about responding on single and concurrent schedules of reinforcement. Fortunately, in this area there are enough data to test both models adequately. First, consider single schedules of reinforcement. It will be convenient to organize the data into three categories, according to the nature of the independent variable being manipulated.

Response Rate as a Function of the Rate of Reinforcement

In the experimental literature data are frequently presented as a relation between the response rate and the obtained rate of reinforcement. It has been observed that this relation is hyperbolic in a surprisingly large number of contexts (de Villiers & Herrnstein, 1976):

$$b_1 = \frac{\alpha r_1}{r_1 + \beta}, \tag{39}$$

where α and β are constants estimated from the data. Herrnstein (1970) showed how the standard assumptions of matching behavior, augmented by the assumption that background reinforcement is noncontingently delivered ($r_i = s_i$ for $i \geqslant 2$), yield the same hyperbolic relation:

$$b_1 = \frac{kr_1}{r_1 + \sum_{i=2}^{n} r_i} \tag{40}$$

$$= \frac{kr_1}{r_1 + \sum_{i=2}^{n} s_i}. \tag{41}$$

In spite of their mathematical similarity, Equations 39, 40, and 41 are very different objects. Equation 39 is an empirical fact or reasonable approximation thereof; Equation 41 is a particular theoretical interpretation of the fact, that is, that the subject is matching and that the estimated background reinforcement is constant; finally, Equation 40 is simply an alternative definition of matching behavior and as such it cannot be tested in a single-schedule experiment.

Now the assumption that background reinforcement comes free or, in the terminology used here, that the background schedules of reinforcement are equivalent to time schedules, is, in a way, very restrictive. One may inquire at this point whether this assumption is really necessary for a matching-based derivation of Equation 39 and, second, whether there exists a maximization-based derivation of the same equation. According to Theorem 6, the surprising answer to the first question is that matching will always lead to Equation 39, provided the feedback function system is hyperbolic. Theorem 6, rewritten here for convenience,

$$b_1 = \frac{\alpha_1^{(m)} r_1}{r_1 + \beta_1^{(m)}},$$

differs from Equation 39 only insofar as the parameters α and β have now acquired a particular interpretation. The constancy of background reinforcement is not necessary for the hyperbolic relationship to hold. In fact, this relationship will be indifferent with respect to the number of background sources of reinforcement; the feedback function for each background source, provided it is approximately hyperbolic; the feedback function for r_1, provided, again, it is approximately hyperbolic; and hyperbolic rescaling of the independent variable (r_1).

Turning now to maximization, one can only state a much weaker result. If s_1 from Equation 5 is substituted into the maximizing response function (Equation 37), then the latter can be solved for b_1, yielding

$$b_1 = \frac{(k + \sum_{i=2}^{n} a_i s_i) r_1}{r_1 + (\sum_{i=2}^{n} a_i^{1/2} s_i) a_1^{-1/2}}$$

or

$$b_1 = \frac{\alpha^{(o)} r_1}{r_1 + \beta^{(o)} a_1^{-1/2}}. \qquad (42)$$

From this equation we see that the hyperbolic relationship can result from maximization in all experiments except those in which the response requirement for the schedule controlling r_1 is itself a variable. Thus, for example, on a VI schedule, where the response requirement does not interact with the received rate of reinforcement, both matching and maximizing would yield a hyperbolic scatter of data points. On VR schedules, however, only matching is consistent with a hyperbola; if we substitute $a_1{}^{1/2} = (b_1/r_1)^{1/2}$ into Equation 42 and solve for b_1, we see that maximization results in a different function:

$$b_1 = \left(\sqrt{\frac{(\beta^{(o)})^2}{4r_1} + \alpha^{(o)}} - \frac{\beta^{(o)}}{\sqrt{4r_1}} \right)^2.$$

As this function is also monotonic, increasing in r_1, it does not offer a qualitative test between the two response rules.

Response Rate as a Function of the Time Requirement

I shall now examine how response rate depends on the response requirement of the RFF. It will be most convenient to use the response functions as they are defined by Equations 34 and 37. Each response function is parametrized by two free parameters, $\alpha^{(m)}$ and $\beta^{(m)}$ or $\alpha^{(o)}$ and $\beta^{(o)}$.

The schedule that most directly manipulates the time requirement is the VI schedule. Depending on the VI feedback function model, the mean scheduled interreinforcement delay in the VI is either identical to the time requirement or, as in the Heyman-Luce (1979a) model, slightly different from it (it is shorter by a factor of $V/[V + I]$). One can thus treat V as being equal to the time requirement d, and its reciprocal $1/V$, which is also the programmed rate of reinforcement, as equal to s. Using the feedback function defined in Equation 18, and substituting into the response functions, yields one set of matching and maximizing VI predictions:

$$b_1^{(m)} = \frac{(\alpha_1^{(m)} - \beta_1^{(m)} kI) V^{-1}}{\beta_1^{(m)} + V^{-1}} \qquad (43)$$

$$b_1^{(o)} = \frac{[\alpha_1^{(o)} - \beta_1^{(o)} (kI)^{1/2}] V^{-1}}{\beta_1^{(o)} (kI)^{1/2} + V^{-1}}. \qquad (44)$$

If one assumes random responding (Equation 12), then the VI response functions are a bit simpler:

$$b_1^{(m)} = \frac{(\alpha_1^{(m)} - \beta_1^{(m)}) V^{-1}}{\beta_1^{(m)} + V^{-1}} \qquad (45)$$

$$b_1^{(o)} = \frac{(\alpha_1^{(o)} - \beta_1^{(o)}) V^{-1}}{\beta_1^{(o)} + V^{-1}}. \qquad (46)$$

The two functions are hyperbolic in $1/V$ for both models; hence, the hyperbolic relation between response rate and the programmed rate of reinforcement should be robust with respect to different assumptions about the IRT distribution.

There are many data that support the hyperbolic VI response function. Figure 2 reproduces the data from a well-known experiment by Catania and Reynolds (1968); Figure 3 is from an experiment by Bradshaw, Szabadi, and Bevan (1976) that is particularly interesting because they used human subjects.

In the absence of independent information about the values of the parameters α^i and β^i, it may seem impossible to differentiate between matching and maximizing single-schedule VI performance. Both are hyperbolic in V^{-1}. This, however, is a seriously misleading conclusion. It is, in fact, very unlikely that a maximizing subject will generate typical VI data. The source of the difficulty lies in the extremely small marginal returns that the VI schedule offers over most of the response range. A maximizing subject distributes his behavior among various alternatives so as to equalize the marginal returns per response over all alternatives.

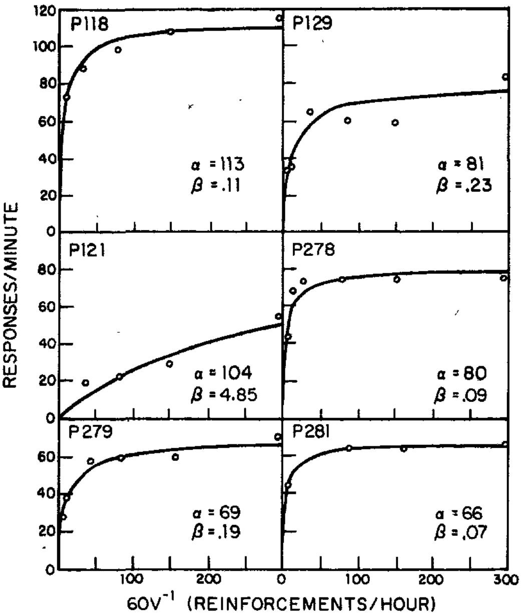

Figure 2. Response rate on variable-interval schedules as a function of the scheduled rate of reinforcement (Catania & Reynolds, 1968). The equation fitted is $b = \alpha V^{-1}/(\beta + V^{-1})$. The estimates of α and β are taken from de Villiers (1977), Figure 8.

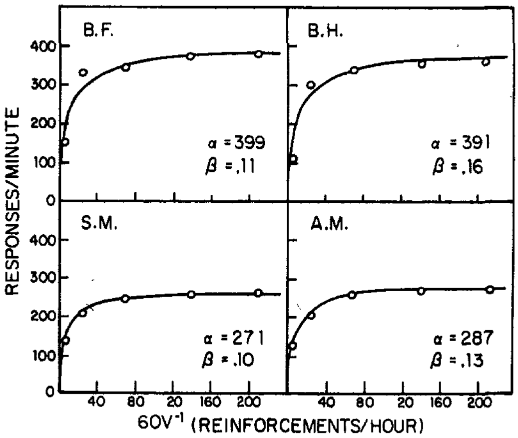

Figure 3. Responding by human subjects on variable-interval schedules as a function of the scheduled rate of reinforcement. (The equation fitted is the same as in Figure 2. Estimates of α and β are taken from Bradshaw, Szabadi, & Bevan, 1976, Table 1.)

To appreciate how small the marginal rate of reinforcement on VI schedules can get, consider Bird 118 in the Catania and Reynolds (1968) study, which responded at a rate of 72 responses/minute on a VI 5.4-minute schedule. Using the random responding feedback function, this implies that the marginal rate of reinforcement at that point was (d = 5.4 minutes/reinforcement; a = 1 response/reinforcement; b = 72 responses/minute)

$$\frac{dr}{db} = \frac{a}{(db + a)^2} < \frac{1}{150{,}000}.$$

On the margin, therefore, the pigeon was exchanging more than 150,000 responses for one instance of reinforcement. Why is this disturbing? Could it not be that the background reinforcement sources were just as weak, that an extra response/minute diverted from the schedule would increase the background reinforcement rate by only 1/150,000 reinforcements per minute? A bit of speculation ought to convince one that this cannot be the case. If it were in fact true that the background marginal rate of reinforcement was so weak, then consider what would happen if the VI schedule were suddenly replaced by a variable-ratio schedule with a response requirement of 150,000 responses per reinforcement. From the point of view of the maximizing subject, this operation leaves the relevant contingencies completely unchanged. The marginal rate of reinforcement on the new schedule is still 1/150,000—now it is fixed at this value regardless of the rate of responding—and it still balances the marginal rate for background reinforcement, presumed also to be 1/150,000. From an elementary application of marginal analysis it follows, therefore, that the subject who was responding 72 times a minute on a VI 5.4-minute schedule should do the same on a VR 150,000.

It is worth noting that several obvious objections to this sort of reasoning make the predicament not less but more severe for maximization. First, one may suggest that the absolute rate of reinforcement influences the effective marginal rate by altering the temporary deprivation level of the subject. But then the marginal rate of reinforcement on the VR 150,000 would be higher, unless one is willing to entertain the counterintuitive notion of increasing marginal value of reinforcement. Second, VI and VR responses may differentially enter into the response constraint. If this is indeed true, then most of the data suggests that VI responses are more "expensive," that is, that a decrease of one VI response can be traded for somewhat more than an increase of one VR response and that, consequently, the same marginal rate of reinforcement should generate higher responding on the VR.[6]

Response Rate as a Function of the Ratio Requirement

The matching and maximizing response functions predict different behavior as the ratio requirement of a schedule is varied. If the subject matches, then by Equation 34 his or her response rate should decrease linearly with the ratio requirement. When the schedule is a pure ratio schedule ($d_1 = 0$), then the response function becomes particularly simple:[7]

$$b_1 = \alpha_1^{(m)} - \beta_1^{(m)} a_1 . \qquad (47)$$

The maximizing response function (Equation 37) is nonmonotonic in the ratio requirement, except when the schedule is a pure ratio schedule. In general, as the ratio requirement is increased from zero, the response rate first also increases from zero, attains a maximum when the ratio requirement is equal to

$$(\sqrt{d^2\beta^{(o)2} + d\alpha^{(o)}} - d\beta^{(o)})^2,$$

and then decreases to zero again. The value of the ratio requirement that induces the highest maximum rate of responding is not a constant but depends, rather, on the value of the time requirement associated with the schedule: The smaller the time requirement, the smaller, also, is this ratio value. In the limiting case when the time requirement is zero, that is, when the schedule is a pure ratio schedule, the response function becomes monotonic, decreasing in the ratio requirement:

$$b_1 = \alpha_1^{(o)} - \beta_1^{(o)} a_1{}^{1/2} \qquad (48)$$

or, more specifically, linearly decreasing with the square root of the ratio requirement.

For pure ratio schedules, then, matching and maximizing predict the same performance in a qualitative sense. Few studies have tested this relation in exactly the way one would like it to be tested: with a VR schedule, a zero-level baseline operant, a large span of ratio requirements, and so forth.

The data in Figure 4 come from an experiment by Zuriff (1970) on multiple VI–VR schedules. The points plotted are the response rates in the VR component; at all the points the VI schedule in the other component was held constant. Matching predicts a linear relation in the left panel, where response rate is plotted against the ratio re-

---

[6] The relevant evidence is summarized by Herrnstein and Heyman (1979).

[7] It is important to mention that the monotonicity of many of the functions derived in this section will not hold if the duration of the reinforcing events is included as part of total session time (as is done, for example, in Staddon, 1979). In particular, the matching response function for ratio schedules (Eq. 47) becomes bitonic in the ratio requirement:

$$b_1 = \frac{a_1(\alpha_1^{(m)} - \beta_1^{(m)} a_1)}{\alpha_1^{(m)} u + a_1(1 - \beta_1^{(m)} u)},$$

where, following Staddon (1979), u is the duration of one reinforcing event.

quirement; maximization predicts the same in the right panel, where response rate is plotted against the square root of the ratio requirement. It seems that the fit is adequate for both panels. On the average, the least squares fit in the matching panel accounts for 83% of the variance, whereas the fit in the maximizing panel accounts for 85% of the variance.

Nevertheless, maximizing provides a somewhat less satisfactory account of VR performance than matching does because the values of the parameters $\alpha^{(o)}$ and $\beta^{(o)}$, which identify the background reinforcement contingencies and the constraint on the total amount of behavior, are grossly different from those estimated from VI data. This problem has been mentioned in the discussion of single-schedule VI performance; it can now be analyzed more precisely.

A fit of Equation 39 to the empirical relation between rate of response and rate of obtained reinforcement on VI schedules will yield estimates of α and β; these in turn can be used to estimate the four key parameters: $\alpha^{(m)}$, $\beta^{(m)}$, $\alpha^{(o)}$, $\beta^{(o)}$. From Equations 39 and 42 and from Theorem 6 it is evident that $\alpha = \alpha^{(m)} = \alpha^{(o)}$ and that $\beta = \beta^{(m)} = \beta^{(o)}a_1^{-1/2}$, where a_1 is the response requirement on the VI schedule. The value of a_1 is not known exactly, but it is almost certainly greater than 1 and less than 10—most likely it is closer to 1 than to 10. Let us set it at 3, with the understanding that it could be off by half an order of magnitude either way. Taking the Catania and Reynolds (1968) data as representative, we see that the median values for α and β among the six subjects (the estimates are taken from de Villiers, 1977) are approximately 80 responses/minute and .15 reinforcements/minute, respectively. This gives an $\alpha^{(m)}$ and $\alpha^{(o)}$ of 80, a $\beta^{(m)}$ of .15, and a $\beta^{(o)}$ of .25. With these numbers in hand, one can make some predictions about how these same subjects should perform on VR schedules. For example, from Equation 47 it follows that if the subject matches, then the value of the ratio requirement at which responding ceases is given by the quantity $\alpha^{(m)}/\beta^{(m)}$. The numerical value for this is 80/.15 or approximately 500 responses—a fairly reasonable number. If the subject is maximizing, however, then Equation 48 implies that responding on the VR should extinguish for ratio values greater than $(\alpha^{(o)}/\beta^{(o)})^2$, and this quantity, for the Catania and Reynolds (1968) subjects, is $(80/.25)^2$ or approximately 100,000 responses—clearly not a reasonable number.

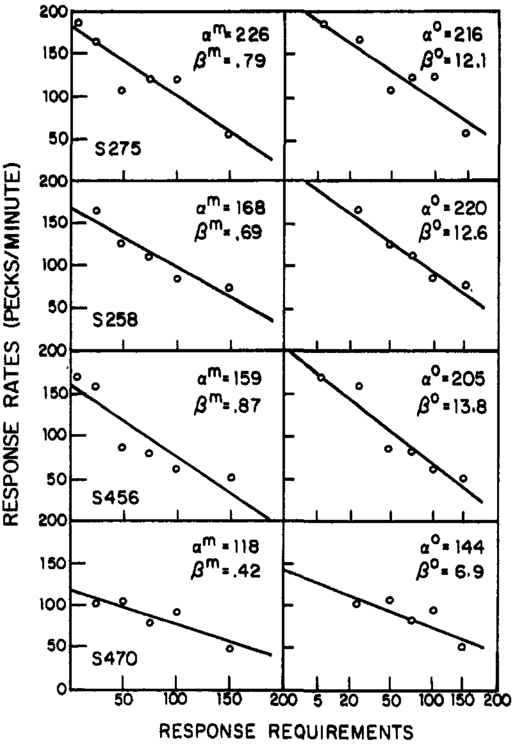

Figure 4. Response rate in the variable-ratio (VR) component of a multiple variable-interval–variable-ratio schedule of reinforcement as a function of the average response requirement in the VR component (Zuriff, 1970). (Equations 47 and 48 are fitted in the left and the right panels, respectively.)

To summarize the argument so far, it seems that if VI and VR single-schedule performances are considered individually, then both the matching and maximizing response rules are compatible with existing data. If the two sets of results are analyzed simultaneously, then the maximizing account breaks down.

Concurrent-Schedule Data

Concurrent schedules of reinforcement provide a more direct opportunity for evaluating the two response rules. Let us first consider concurrent VI–VI schedules.

The Concurrent VI–VI Experiment

The question that has most often been asked about this experiment is what the relationship is between the response ratio b_1/b_2, the obtained rate of reinforcement ratio r_1/r_2, and the scheduled rate of reinforcement ratio V_2/V_1. The answer is that, to a first approximation at least, these quantities should all be equal regardless of whether the subject matches or maximizes. If the subject matches, then the equality between the response and the obtained rate of reinforcement ratios will hold by definition; the other equalities depend crucially on the assumptions that the response requirements for the two schedules are the same and that the time requirements are equal to the mean schedule delays V_i. Both of these assumptions are either exactly true or only slightly inaccurate for all the VI schedule feedback functions we have here considered. The equality of the response ratio and the scheduled-rate-of-reinforcement ratio is, for both matching and maximizing, a consequence of Theorem 1 (see Equation 26); the maximization-based equality between the response ratio and the received rate of reinforcement ratio is a consequence of Theorem 2.

Because both matching and maximizing lead to the above equalities, it appears that the concurrent VI–VI experiment is a singularly unsuitable vehicle for a qualitative test of the two response rules. Nevertheless, the historical importance of this experiment and the relative abundance of data about it have made it a choice object of theoretical argument. Instead of commenting on this debate, which revolves about some important but fine points of the VI feedback function (see Heyman, 1979; Heyman & Luce, 1979; Rachlin, 1979), I will discuss an altogether different aspect of the concurrent VI–VI experiment, which rules out maximization as the response rule guiding behavior on these schedules.

First, I will simplify the problem by assuming that the subject has already allocated an optimal quantity of behavior to background schedules and that, in order to obtain the maximum reinforcement, it is only necessary for the subject to divide optimally the remaining behavior k between the two VI schedules. In other words, background reinforcement does not enter into the problem. To allocate responding optimally between the VI schedules, the subject then has to maximize the expression

$$r_1 + r_2 = \frac{b_1}{V_1 b_1 + kI} + \frac{k - b_1}{V_2(k - b_1) + kI} \quad (49)$$

or, in the Staddon and Motheral (1978) model,

$$r_1 + r_2 = \frac{b_1}{V_1 b_1 + 1} + \frac{k - b_1}{V_2(k - b_1) + 1}. \quad (50)$$

In theory, as Staddon and Motheral have shown, this function, as well as the previous one, has a maximum that is located at the matching-behavior distribution: $b_1/(k - b_1) = r_1/r_2$. In reality, however, when reasonable values for the parameters V_1, V_2, k, and I are inserted, this function is so flat that one cannot claim that a meaningful maximum exists at all (see Heyman & Luce, 1979a, Figure 4). For example, in a concurrent VI-3-minute: VI-6-minute experiment, the maximum rate of reinforcement is obtained when the subject responds twice as much on the VI 3-minute as on the VI 6-minute schedule. If the subject's total response rate is 100 responses/minute, and if the response requirement—kI—is three responses, then the maximum rate of reinforcement is 29.56 reinforcements/hour. However, indifference between the two schedules will result in 29.51 reinforcements/hour—a decrease of only .2%; even a 2:1 preference for the poorer schedule will still yield 99.3% of the maximum rate.

When the Staddon and Motheral (1978) equation is used ($kI = 1$), the flatness becomes even more pronounced. In my example the subject would collect at least 99% of the maximum rate as long as he or she distributed responding somewhere between a 7:1 preference for the VI-6′ and a 20:1 preference for the VI-3′. What Staddon and Motheral have really shown is that the subject who maximizes Equation 50 will in many concurrent VI–VI experiments be lost; that is, the subject's behavior distribution will not be determined by the schedules of reinforcement but by chance. The purely random fluctuations in reinforcement rate from session to session are several times greater than the fluctuations caused by different response distributions. The formidable nature of the problem of differentiating a 29.56 reinforcement rate from a 29.51 or a 29.34 rate emerges very clearly when it is reformulated as a signal-detection task. Given that reinforcement delivery is approximately Poisson, the standard deviation of the received number of reinforcements over any time period is equal to the square root of the expected received number. Over a 1-hour session, therefore, the standard deviation of about five reinforcements will surely overwhelm the minute differences caused by the different response distributions. Even with very keen statistical know-how, subjects would need hundreds of hours of sampling at the three distributions (2:1, 1:1, or 1:2 response ratios) to be able to distinguish their payoff rates. If the maximizing subject's behavior is controlled by an estimate of total reinforcement rate that is based on a few hours of session time, then, in concurrent VI–VI experiments, the maximum for all practical purposes doesn't exist. If, on the other hand, the estimate is based on a large number of session hours, then a maximum may eventually emerge; however, the subject will in that case take an extremely long time in settling down to a stable response pattern. The wide scatter of resulting behavior distributions across subjects and the drift from day to day that one should see in individual subjects are not commonly found in concurrent VI–VI experiments. In fact, many animals start to match very quickly, sometimes within a session or two.

In contrast to maximization, a matching response rule requires but the crudest estimators of reinforcement rate. After 1 hour of session time, a VI 6-minute schedule will deliver 10 ± 3 reinforcements, and a VI 3-minute schedule, 20 ± 4 reinforcements; the chances for substantially confusing the two rates with any kind of estimators are remote.

The Concurrent VI–VR Experiment

As it leads to numerous qualitative differences between matching and maximizing behavior, the concurrent VI–VR experiment is the most useful of the ones considered here. At the root of these differences is one counterintuitive consequence of matching behavior that may be phrased as follows: If there is a ratio schedule in the environment that supports a nonzero rate of behavior, then the total rate of reinforcement received from all the schedules in the environment will be constant irrespective of the other nonratio schedules. Perhaps the quickest way to demonstrate the truth of this statement is to think of the total amount of responding as a single response class that satisfies the preconditions for Theorem 4(M). It then follows from the theorem that the received rate of the reinforcement—in this case also the total rate of reinforcement—is a linear function of the rate of response, which, in this case, always has to equal k. The constancy of the total rate of reinforcement in such an environment does not, incidentally, depend on the hyperbolic feedback function system. For if each r_i is some arbitrary function f_i of the rate of response b_i, and if r_1 is a linear function of b_1, it still follows from matching that $r_i/b_i = r_1/b_1$ or $f_i(b_i)/b_i = f_1(b_1)/b_1$, implying that $f_i(b_i) = b_i f_1(b_1)/b_1 = b_i b_1 a_1{}^{-1}/b_1 = b_i/a_1$. The total rate of reinforcement is then equal to k/a_1 as

$$\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} f_i(b_i) = \sum_{i=1}^{n} b_i/a_1 = k/a_1 .$$

The behavioral implications of this matching result can be quite unusual. Most significantly, subjects will not be able to take advantage of favorable changes in the feedback function system, changes that should allow them to extract a higher rate of reinforcement. To any change in the feedback system they will respond by redistributing their behavior so as to maintain the old total reinforcement rate, not so as to collect additional reinforcements from the new sources. If the extra reinforcement is free, for example, it will draw a proportional amount of behavior away from the ratio schedule, thereby canceling out the gain with an equal loss (Prelec & Herrnstein, 1978). Conversely, a change in the opposite direction, say a reduction in free reinforcement, will not reduce their total reinforcement because their original behavior distribution was not an optimal one.

The mathematics of this reasoning are not altered if one substitutes the language of utility or value for the language of reinforcement. Insofar as the total rate of reinforcement corresponds to the utility level of the organism, one may conclude that the matching subject, when placed in an environment that includes a ratio schedule, behaves as if his or her sole principle of action was the maintenance of a constant utility level. The subject symmetrically resists increases and decreases in that level.

Now let us look at some data. Herrnstein and Heyman (1979) performed the relevant experiment with concurrent VI–VR schedules and found that their subjects, by not distributing their responses in the optimal manner, gave up, on the average, 60 reinforcements per hour, or 30% of their total rate. At the same time, the subjects matched, but with an overall response-rate bias of 1.8 for the ratio schedule.[8] From the Herrnstein and Heyman data it is also possible to check the claim—which is not affected by the presence of bias—that the total rate of reinforcement should be constant as long as there is an active ratio schedule in the environment. Among the various VR–VI combinations that they ran, three had the same VR schedule coupled with a different VI schedule. Specifically, the ratio schedule was always a VR-30, whereas the time requirement on the concurrent VI was either 40, 30, or 5 sec. The scheduled rate of reinforcement on the VI thus ranged from 90 to 240 reinforcements per hour. Of the four subjects in the experiment, two responded very little or not at all on the VR (5% and 15%) so that their total rate of obtained reinforcement naturally increased along with the scheduled rate on the VI. The other two subjects responded more substantially on the VR (37% and 61%) and, as predicted, their total rate of reinforcement was virtually constant for the three conditions. When the scheduled rate of reinforcement on the VI was 90 reinforcements/hour, then their average obtained rate of reinforcement from both schedules was 269 reinforcements/hour; when the VI scheduled rate was 120 reinforcements/hour, their total received rate decreased slightly to 249 reinforcements/hour; finally, when the VI scheduled rate was 240 reinforcements/hour, the received rate climbed back up to 271 reinforcements/hour. Even without a quantitative argument, one can tell that this is nonoptimal behavior. The difference in the scheduled rate of reinforcement between the two extreme conditions is 150 reinforcements/hour; simply maintaining the same pattern of responding in the VI 15-sec schedule as they had in the VI 40-sec schedule would have allowed the subjects to collect a substantial number of the extra 150 reinforcements. Characteristically, they responded too much on the VI and too little on the VR—the same sort of error from the point of view of maximization that has been detected in the analysis of single-schedule VI and VR performance. This points very strongly to the conclusion that the nonoptimalities of operant behavior have a common source: too much responding for the small marginal returns that VI schedules offer.

Implications for General Theories of Behavioral Allocation

The two theories developed here have relied on a number of simplifying assumptions, namely, that reinforcement values are constant, that operant responding is not intrinsically rewarding or punishing, and that each RFF has the shape of a hyperbola. The two theories, therefore, are viewed as special cases, or prototypes, of the more general theories of behavioral allocation (e.g., Rachlin & Burkhard, 1978; Staddon, 1979) that attempt to explain how time and effort are divided across disparate activities in the presence of arbitrary constraints. In this section I will assess the implications of my comparison of matching and maximizing for the general theories of behavioral allocation.

The first question that comes up in a competitive evaluation of prototypes is the ques-

---

[8] I interpret this bias as due to a difference in *k* for VI- and VR-controlled pecking. (See Figures 2 and 4.) The response-rate ceiling seems to be about twice as high for ratio-maintained performance (Figure 4). In other words, the pigeon who is pecking 100 times per minute on a VI is doing nothing but pecking, whereas the pigeon who pecks at that same rate on the VR schedule is investing half of his behavior elsewhere.

tion of fairness: Do the simplifying assumptions bear equally hard on both models? (Whether the simplifying assumptions are literally true is much less important.) To ensure fairness the two theories were developed in a parallel way. Not only do they have the same number of free parameters but their formal structures are in all other aspects remarkably similar. As Theorems 1 and 2 show, what separates the predictions of one theory from those of the other is a transformation of the parameter space. Matching in one experiment is maximizing in another. Hence, the greater ease with which the matching response functions fit the data cannot be attributed to any asymmetry between the two theories.

The second question is that of representativeness: Does the failure of a prototype entail the failure of the full-scale model? Given that the simplifying assumptions are really boundary conditions that single out these theories as special cases of the general theories, will a more flexible maximizing theory, by suitably adjusting the boundary conditions, be able to explain that which the special one could not?

To show that some difficulties persist even at a higher level of abstraction, I will examine a general maximizing theory developed by Rachlin and Burkhard (1978) that is based on the notion of maximization of value. The subject, according to this theory, is engaged in allocating time among various activities. Different time allocations have different values associated with them, and the fundamental principle guiding the subject's choice among the different available time allocations is to act so as to maximize the value of the chosen allocation. Schedules of reinforcement control performance by restricting the set of possible time allocations, thus making some previously attainable allocations impossible. Operant behavior is a function of two independently specifiable elements: an ordinal preference structure over the set of all time allocations (expressed by a value function) and the geometry of the set of time allocations permitted by the schedule (the feasible set).

With a modest amount of translation, one can see how this theory would attempt to analyze a concurrent-schedule experiment. A time allocation is described by three quantities: the time spent consuming reinforcement (C), the time spent responding (I), and the time spent doing anything else (N). It is standard to assume (e.g., Rachlin, 1978) that C and I are proportional to the rates of reinforcement and response:

$$\begin{aligned} C &= \gamma_f(r_1 + r_2) \\ I &= \gamma_r(b_1 + b_2), \end{aligned} \tag{51}$$

where γ_f and γ_r are the average durations of one reinforcement and one response, respectively. The subject, it is assumed, is maximizing a continuous, differentiable, downward concave function $V(C, I, N)$ subject to two additional constraints. First, the rates of reinforcement are dependent on the rates of response via the two RFFs, whose exact shape for the moment we ignore:

$$\begin{aligned} r_1 &= f_1(b_1) \\ r_2 &= f_2(b_2). \end{aligned} \tag{52}$$

Second, total time is equal to session time T:

$$I + C + N = T. \tag{53}$$

Because a pair of response rates uniquely determines a time allocation, through the equalities

$$I = \gamma_r(b_1 + b_2), \tag{54}$$

$$C = \gamma_f[f_1(b_1) + f_2(b_2)], \tag{55}$$

$$\begin{aligned} N = \mathrm{T} &- \gamma_r(b_1 + b_2) \\ &- \gamma_f[f_1(b_1) + f_2(b_2)], \end{aligned} \tag{56}$$

it is possible to solve the maximizing problem by differentiating V with respect to b_1 and b_2 and letting the derivatives equal zero:

$$\begin{aligned} \frac{\partial V}{\partial b_1} &= \gamma_r \frac{\partial V}{\partial I} + \gamma_f \frac{\partial V}{\partial C}\frac{df_1}{db_1} \\ &\quad - \frac{\partial V}{\partial N}\left(\gamma_r + \gamma_f \frac{df_1}{db_1}\right) = 0 \\ \frac{\partial V}{\partial b_2} &= \gamma_r \frac{\partial V}{\partial I} + \gamma_f \frac{\partial V}{\partial C}\frac{df_2}{db_2} \\ &\quad - \frac{\partial V}{\partial N}\left(\gamma_r + \gamma_f \frac{df_2}{db_2}\right) = 0. \end{aligned} \tag{57}$$

A necessary condition for maximum is that $\partial V/\partial b_1 = \partial V/\partial b_2$ and, by inspection of the above pair of equations, this condition yields the simple equality

$$\frac{df_1}{db_1} = \frac{df_2}{db_2}. \qquad (58)$$

The marginal substitutabilities defined by the value function—terms such as $(\partial V/\partial C)/(\partial V/\partial I)$—cancel out, leaving nothing but a restatement of Assumption 5(O)′, the equal marginal-rate-of-reinforcement rule. Without knowing the value function itself we cannot predict the total response level $(b_1 + b_2)$, but given any total response level, we can predict, irrespective of the value function, how it should be divided between b_1 and b_2. So all of the arguments previously raised against maximization in concurrent-schedule experiments apply here with equal force: On concurrent VI–VR schedules the equal marginal-rate rule is violated outright (Herrnstein & Heyman, 1979); on concurrent VI–VI schedules the marginal rates are too small to define meaningfully a maximum at the observed, that is, matching, response distribution.[9]

Rachlin et al. (1980) have suggested that the failure to maximize in concurrent VI–VR experiments might be related to a difference in the amount of leisure that the two schedules provide: "A ratio schedule provides less leisure activity for the same amount of food reinforcement. . . . To the extent that leisure is actually substituted for food, one should find a bias toward the interval schedule" (p. 370).

Perhaps so, but an explanation along these lines cannot be readily translated into the formalism of value-maximization theory. There are only two ways in which the theory could be made more general than it is already: (a) It could deal with a longer list of activities, and (b) it could contain additional feedback function constraints among them. Clearly (a) is not intended by Rachlin et al. (1980) because in that same paragraph they define *leisure* as "activity other than key pecking or eating" (p. 370). But if we stay with a threefold activity classification, then there is no room for an additional response–leisure feedback function—something like $N = h_1(b_1) + h_2(b_2)$—because any pair of response rates already uniquely determines leisure time via Equation 56.

What, then, could have been intended by their explanation? The intuition behind it, I believe, is that the subject counts separately leisure and food obtained from the two schedules, as N_1, N_2 and C_1, C_2, and is then influenced in some way by the relative magnitudes N_1/C_1 and N_2/C_2. But variables of this kind, for example, N_1, N_2, C_1, C_2, have no role to play within a maximization theory; the mathematics of value maximization wipes out the distinction between N_1 and N_2, between C_1 and C_2, and, for that matter, between I_1 and I_2, as long as the respective pairs of activities are perfect substitutes. It would be entirely correct to say that value-maximizing subjects do not know how leisure, food, and responding are divided be-

---

[9] One could quarrel with the first part of Equation 51 by claiming that a given rate of response on a VI schedule indicates more time spent responding than would be predicted by that same rate on a VR. The model can be remedied as follows: Assume that Schedule 1 is the VI and Schedule 2 the VR. Total time spent responding should then be given by

$$I = \gamma_1 b_1 + \gamma_2 b_2,$$

where γ_1 is the coefficient of proportionality for the VI and γ_2 for the VR schedule, respectively. Note that $\gamma_1 > \gamma_2$, by assumption. Following through on the derivation in Equations 51–58, one arrives at a more complicated version of Equation 58:

$$\frac{df_2}{db_2} = \frac{df_1}{db_1} + \left[\frac{\gamma_1 - \gamma_2}{\gamma_f}\right]\left[\frac{(\partial V/\partial I) - (\partial V/\partial N)}{(\partial V/\partial C) - (\partial V/\partial N)}\right].$$

In this case, the marginal values did not cancel out, and the equal marginal-rate-of-reinforcement rule no longer applies. Does this rescue value maximization? No, because the only plausible ordering of the marginal values of C, I, and N is

$$\partial V/\partial C > \partial V/\partial N > \partial V/\partial I,$$

that is, the marginal value of 1 sec of food is greater than the marginal value of 1 sec of leisure, which is greater than the marginal value of 1 sec of work. Checking the signs, and remembering that $\gamma_1 > \gamma_2$, we see that the value maximization implies

$$\frac{df_2}{db_2} < \frac{df_1}{db_1},$$

or that the marginal rate of reinforcement on VR schedules has to be smaller by some unspecified amount than the marginal rate of reinforcement on VI schedules. This is even less realistic than Equation 58.

tween the two schedules. That, for them, is an idle question.

A value function that had separate entries for VI leisure time and VR leisure time, thus treating the two as completely different activities with different marginal substitutabilities, could, indeed, predict violation of Equation 58. But this expansion of the value function would be a most costly maneuver, for it would imply that the shape of the feasible set and the underlying preference order were not independent theoretical elements, both, in this case, being dependent on the schedule of reinforcement. And without independence little remains of the notion of value.

How Important Is the Hyperbolic RFF Assumption?

Most of the difficulties with the maximizing account of equilibrium behavior can be traced to the VI schedule and its vanishing marginal returns. As different assumptions about the shape of the RFF produce different estimates of these marginal returns, it is not unreasonable to ask whether some alternative VI feedback function may attenuate the discrepancy between the VI and VR marginal reinforcement rates and thus lead to predictions more congenial to maximization theory. Rachlin and his associates (Rachlin et al., 1980), in particular, have argued for a power function RFF space. They do agree, however, that

> no single feedback function can exactly model all laboratory procedures . . . The fact that even with fairly drastic procedural variations (hence, different feedback functions) behavior with individual and concurrent variable-interval schedules does not vary greatly indicates that precise specification of the feedback function is not necessary. (p. 362)

Of the various procedural variations that have been tried so far, one deserves to be mentioned here, for it demonstrates convincingly that the exact form of the VI-schedule RFF has no bearing on the maximization question. Vaughan (1976) has developed linear VI schedules that differ from ordinary VI schedules in that they make the obtained rate of reinforcement totally insensitive to fluctuations in the response rate over the observed ranges of fluctuation. A linear VI 1-minute schedule, for example, will deliver 60 reinforcements per hour as long as the subject responds at least 60 times during the hour. As far as one can tell, behavior maintained by these schedules looks no different from behavior maintained by ordinary VI schedules. For the six birds in Vaughan's study, the average rate of response maintained by linear VI 1-minute schedules was 65 responses per minute. This is almost exactly what one would predict from the Catania and Reynolds (1968) data in Figure 3. Nor are there any noticeable differences in concurrent-schedule behavior with linear VIs (Herrnstein & Vaughan, 1980). One can take it as an empirical fact, therefore, that making the marginal rate of reinforcement exactly equal to zero does not grossly change VI performance.

With this in mind, the challenge to maximization theory can be formulated even on the basis of single-schedule linear VI performance. Why does the subject respond so much? The special theory of reinforcement maximization does not have an answer, but the general theory can provide one along the following lines: (a) When the linear VI schedule is introduced, it fixes the level of the contingent response (eating) and so moves the subject into a new region of time allocation space; (b) in this new region the marginal substitutabilities of various responses are different from what they were previously; (c) the marginal value of the operant (instrumental) response, in particular, is higher relative to the marginal value of leisure; (d) this pushes operant responding up at the expense of leisure until a new equilibrium between their marginal values is established; (e) nothing in (a) through (d) precludes a high level of operant responding at the new equilibrium point.

But consider what should happen if one reverses the roles of leisure and pecking in Vaughan's (1976) experiment. Specifically, one now programs the linear VI schedule to reinforce a brief period (Δt seconds) of not pecking (i.e., leisure), and one chooses Δt sufficiently small to ensure that enough "leisure events" occur to collect all 60 reinforcements. What will subjects do when faced with the new procedure? If they are value maximizers in the sense intended by Rachlin and Burkhard (1978) or, for that matter, by Staddon (1979), they will ignore this change and maintain the old response distribution. The only effective constraint imposed by the linear VI schedule is a ceiling on total eating time, and this ceiling remains the same. Because the value function, presumably, has also not changed, the solution to the constrained maximization problem will in both conditions be the same.

Faced with this highly counterintuitive prediction, advocates of value maximization can respond in a number of ways. First, they may claim that the threefold activity classification is too coarse and that, for example, a second of leisure followed by food belongs to a different activity than a second of leisure not followed by food. This will create an embarrassing proliferation of activities, with moments of time classified not just by what the subjects are doing but also by what they just did and what they will do.

A more attractive alternative is to allow for a time-discounting process of some kind. The resulting theory will have to be a molecular one, because the present impact of appropriately discounted future consequences needs to be continuously recalculated in time. The similarity here with momentary maximizing is not coincidental, for momentary maximizing is nothing but the extreme example of temporal discounting: Out of the chain of consequences that a response initiates, the subject counts only the immediate one and ignores the rest.

Before making this move, the maximization theorist must appreciate the fact that the relationship between a maximization theory with temporal discounting and a molar maximization theory of the kind proposed by Rachlin and Burkhard (1978) or Staddon (1979) is a truly antagonistic one. The reason why the molecular theory cannot underwrite value maximization is not that the limited time horizon of the subject who discounts future consequence offends one's sense of rationality or one's convictions about what the creature should be maximizing, but rather, that such a subject will not reveal a consistent set of molar preferences. Let us return to the linear VI thought experiment. Surely the myopic subject will peck more in the condition in which the schedule reinforces a peck than in the condition in which the schedule reinforces its absence. On a linear VI schedule, to collect a reinforcement now is to forfeit one later. Any preference for the present one over the future one will show up, on the molar level, as a shift in the marginal substitutability of leisure for pecking between the two conditions. The time-discounting subject will register a molar-level change in preference as a result of a procedural change that the molar theory does not recognize.

Conclusion

At the higher level of abstraction that characterizes the general theories, the conflict between matching and maximizing can be seen as a conflict between two mutually incompatible theories of value (or reward). The difference is simple but fundamental: Maximizing subjects have one value scale that they are trying to maximize; matching subjects have *n* value scales—one for each activity—that they are trying to equalize. Both notions of value are incomplete: Maximizers have no notion of the value of individual activities—they do not rank activities any particular way; matchers have no notion of total value across activities—they do not rank behavioral allocations in any particular way. Maximizers must make intertemporal comparisons so they can tell if they are moving in the direction of a preferred allocation; matchers must compare activities so that they can tell if they are equalizing their values. In sum, the maximizer aggregates value across activities and disaggregates it by time; the matcher disaggregates it between activities and aggregates it through time.

The equilibrium time allocation of the subject is thus subject to two radically different interpretations. In everyday language the question is whether subjects, at equilibrium are equally happy in each of their *n* activities or whether they are maximally happy across the *n* activities.

It is entirely possible that real creatures possess both types of dimensions of value. For example, one can usually tell how satisfied one feels in each activity and also how satisfied one feels across several activities, for example, during a day or a week. Where both dimensions exist, however, there may be conflict. Wishing to be equally happy in each activity one may be less than maximally happy; if one is maximally happy, one should be prepared to tolerate unequal levels of happiness along the way. The more ways one can aggregate and disaggregate the hedonic stream, the more tension there will be between the demands of the alternative measures.

An organism "has" a value structure to the extent that its behavior is controlled by variables required by that structure. The full preference order over behavioral allocations makes contact with performance by generating marginal substitutabilities that then balance the marginal trade-offs allowed by the schedules of reinforcement (the marginal rates of transformation, as the economist would call them). If subjects do not react to these marginal trade-offs, then their behavior cannot be accounted for by value-maximization theory: With a bit of ingenuity one can always exploit insensitivity to marginal trade-offs so as to show a failure of some fundamental preference axiom (e.g., transitivity).

Therefore, the evidence from experiments with single and concurrent schedules of reinforcement calls into question the psychological reality of all theoretical entities derived from a stable preference order over behavioral allocations: value functions, indifference curves, marginal substitutabilites. The subjects' behavior may still be orderly, but it will not be orderly in the way required by value-maximization theory. The extent to which a value structure compatible with the matching response rule—a value structure that interprets behavioral equilibria as points of equal value—can uncover and systematize the orderliness remains to be seen.

References

Baum, W. M. The correlation-based law of effect. *Journal of the Experimental Analysis of Behavior*, 1973, *20*, 137-153.

Baum, W. M. On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior*, 1974, *22*, 231-242.

Baum, W. M. Optimization and the matching law as accounts of instrumental behavior. *Journal of the Experimental Analysis of Behavior*, 1981, *36*, 387-403.

Baum, W. M., & Rachlin, H. C. Choice as time allocation. *Journal of the Experimental Analysis of Behavior*, 1969, *12*, 861-874.

Bradshaw, C. M., Szabadi, E., & Bevan, P. Behavior of humans in variable-interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, 1976, *26*, 135-141.

Catania, A. C., & Reynolds, G. S. A quantitative analysis of the responding maintained by interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, 1968, *11*, 327-383.

de Villiers, P. A. Choice in concurrent schedules and a quantitative formulation of the law of effect. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior*. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

de Villiers, P. A., & Herrnstein, R. J. Toward a law of response strength. *Psychological Bulletin*, 1976, *83*, 1131-1153.

Herrnstein, R. J. Relative and absolute strength of response as a function of the frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, 1961, *4*, 267-272.

Herrnstein, R. J. On the law of effect. *Journal of the Experimental Analysis of Behavior*, 1970, *13*, 243-266.

Herrnstein, R. J. Formal properties of the matching law. *Journal of the Experimental Analysis of Behavior*, 1974, *21*, 159-164.

Herrnstein, R. J., & Heyman, G. M. Is matching compatible with reinforcement maximization on concurrent variable interval, variable ratio? *Journal of the Experimental Analysis of Behavior*, 1979, *31*, 209-223.

Herrnstein, R. J., & Loveland, D. H. Maximizing and matching on concurrent ratio schedules. *Journal of the Experimental Analysis of Behavior*, 1975, *24*, 107-116.

Herrnstein, R. J., & Vaughan, W., Jr. Melioration and behavioral allocation. In J. E. R. Staddon (Ed.), *Limits to action: The allocation of individual behavior*. New York: Academic Press, 1980.

Heyman, G. M. Matching and maximizing on a continuous choice procedure. Unpublished doctoral dissertation, Harvard University, 1977.

Heyman, G. M. A Markov model description of changeover probabilities on concurrent variable-interval schedules. *Journal of the Experimental Analysis of Behavior*, 1979, *31*, 41-51.

Heyman, G. M., & Luce, R. D. Operant matching is not a logical consequence of maximizing reinforcement rate. *Animal Learning and Behavior*, 1979, *7*, 133-140. (a)

Heyman, G. M., & Luce, R. D. Reply to Rachlin. *Animal Learning and Behavior*, 1979, *7*, 268-270. (b)

Intriligator, M. D. *Mathematical optimization and economic theory*. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.

Prelec, D., & Herrnstein, R. J. Feedback functions for reinforcement: A paradigmatic experiment. *Animal Learning and Behavior*, 1978, *6*, 181–186.

Rachlin, H. A molar theory of reinforcement schedules. *Journal of the Experimental Analysis of Behavior*, 1978, *30*, 345–360.

Rachlin, H. Comment on Heyman and Luce. *Animal Learning and Behavior*, 1979, *7*, 267–268.

Rachlin, H., & Burkhard, B. The temporal triangle: Response substitution in instrumental conditioning. *Psychological Review*, 1978, *85*, 22–47.

Rachlin, H., Green, L., Kagel, J. H., & Battalio, R. C. Economic demand theory and psychological studies of choice. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 10). New York: Academic Press, 1976, pp. 129–154.

Rachlin, H. C., Kagel, J., & Battalio, R. C. Substitutibility in time allocation. *Psychological Review*, 1980, *87*, 355–374.

Shimp, C. P. Optimal behavior in free-operant experiments. *Psychological Review*, 1969, *76*, 97–112.

Skinner, B. F. *The behavior of organisms*. New York: Appleton-Century-Crofts, 1938.

Staddon, J. E. R. Operant behavior as adaptation to constraint. *Journal of Experimental Psychology: General*, 1979, *108*, 48–67.

Staddon, J. E. R. Optimality analyses of operant behavior and their relation to optimal foraging. In J. E. R. Staddon (Ed.), *Limits to action: The allocation of individual behavior*. New York: Academic Press, 1980.

Staddon, J. E. R., & Motheral, S. On matching and maximizing in operant choice experiments. *Psychological Review*, 1978, *85*, 436–444.

Vaughan, W., Jr. Optimization and reinforcement. Unpublished doctoral dissertation, Harvard University, 1976.

Zuriff, G. E. A comparison of variable-ratio and variable-interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, 1970, *13*, 369–374.

Appendix A

Derivations for Theorems 1–7

This appendix has three purposes: it provides derivations for the main results of the Basic Properties of Matching and Maximizing Behavior section; it presents, for comparison, several derivations based on the power RFF system; and it shows briefly how one would go about investigating the properties of matching and maximizing behavior with a general system of feedback functions.

Let $\{h_i\}$ be a set of feedback functions so that $r_i = h_i(b_i)$ for $i = 1, \ldots, n$. Each h_i generates two important auxiliary functions, the local RFF $h_i^*(b_i)$ and the marginal RFF $h_i'(b_i)$:

$$h_i^*(b_i) = \frac{h_i(b_i)}{b_i}$$

$$h_i'(b_i) = \frac{dh_i}{db_i}(b_i).$$

For hyperbolic RFFs the local and marginal RFFs are

$$h_i^*(b_i) = (d_i b_i + a_i)^{-1} \qquad \text{(A1)}$$

and

$$h_i'(b_i) = a_i(d_i b_i + a_i)^{-2}. \qquad \text{(A2)}$$

For Rachlin's power RFF, which we write as $r_i = d_i b_i^{a_i}$, they are

$$h_i^*(b_i) = d_i b_i^{a_i - 1} \qquad \text{(A3)}$$

$$h_i'(b_i) = a_i d_i b_i^{a_i - 1}. \qquad \text{(A4)}$$

As was explained in the Assumptions of the Special Theories section, to find the complete matching and maximizing response distributions one has to consider the following systems of equations:

$$\left.\begin{aligned} &\sum_{i=1}^{n} b_i = k, \\ &r_i = h_i(b_i) \quad i = 1, \ldots, n, \\ &h_i^*(b_i) = \lambda \quad i = 1, \ldots, n \end{aligned}\right. \qquad \text{(A5)}$$

and

$$\left.\begin{aligned} &\sum_{i=1}^{n} b_i = k, \\ &r_i = h_i(b_i) \quad i = 1, \ldots, n, \\ &h_i'(b_i) = \lambda \quad i = 1, \ldots, n. \end{aligned}\right. \qquad \text{(A6)}$$

Both Equations A5 and A6 contain $2n + 1$ equations and that is exactly the number needed to solve for the $2n + 1$ unknowns (n response rates, n reinforcement rates, and one constant λ). As in the text, I will label the matching and maximizing equilibrium values of the $2n + 1$ variables with the superscripts (m) and (o), respectively.

Theorem 1: Pairwise Linearity

Directly from Equations A5 and A6 it follows that matching and maximizing imply for any pair of schedules i and j

$$h_j^*(b_j^{(m)}) = h_i^*(b_i^{(m)}), \qquad \text{(A7)}$$

$$h_j'(b_j^{(o)}) = h_i'(b_i^{(o)}), \qquad \text{(A8)}$$

or

$$b_j^{(m)} = h_j^{*\,-1}(h_i^*(b_i^{(m)})), \qquad \text{(A9)}$$

$$b_j^{(o)} = h_j'^{\,-1}(h_i'(b_i^{(o)})). \qquad \text{(A10)}$$

To economize on parentheses, the composition of functions such as $h_j^{*-1}(h_i^*(b_i^{(m)}))$ will be written as $h_j^{*-1}h_i^*(b_i^{(m)})$.

Given matching or maximizing, pairwise linearity will hold if the composed functions $h_j^{*-1}h_i^*$ or $h_j'^{-1}h_i'$ are linear. This is indeed true for hyperbolic RFFs; substituting from Equations A1 and A2 into Equations A7 and A8 gives

$$(d_i b_i^{(m)} + a_i)^{-1} = (d_j b_j^{(m)} + a_j)^{-1} \qquad \text{(A11)}$$

and

$$a_i(d_i b_i^{(o)} + a_i)^{-2} = a_j(d_j b_j^{(o)} + a_j)^{-2}. \qquad \text{(A12)}$$

Solving for $b_i^{(m)}$, $b_i^{(o)}$ proves Theorem 1.

The power function does not imply pairwise linearity but, instead, log-linearity; applying the above argument to Equations A3 and A4 and taking logarithms one obtains

$$\log b_j^{(m)} = (a_j - 1)^{-1} \log (d_i/d_j) + (a_i - 1)/(a_j - 1) \log b_i^{(m)},$$

$$\log b_j^{(o)} = (a_j - 1)^{-1} \log (a_i d_i/a_j d_j) + (a_i - 1)/(a_j - 1) \log b_i^{(o)}.$$

Theorem 5: Response and Demand Functions

Provided that h_i^*, h_i' are invertible, one can always write $b_j^{(m)}$, $b_j^{(o)}$ as functions of $b_i^{(m)}$, $b_i^{(o)}$ via Equations A9 and A10. Summing up over all $b_j^{(m)}$, $b_j^{(o)}$ creates an equation that implicitly defines $b_i^{(m)}$, $b_i^{(o)}$ in terms of the functional parameters and of k:

$$\sum_{j=1}^{n} b_j^{(m)} = \sum_{j=1}^{n} h_j^{*\,-1}h_i^*(b_i^{(m)}) = k, \qquad \text{(A13)}$$

$$\sum_{j=1}^{n} b_j^{(o)} = \sum_{j=1}^{n} h_j'^{\,-1}h_i'(b_i^{(o)}) = k. \qquad \text{(A14)}$$

These equations may or may not be solvable for $b_i^{(m)}$, $b_i^{(o)}$. With hyperbolic RFFs they are solvable, because each $h_j^{*-1}h_i^*$ and $h_j'^{-1}h_i'$ is linear in b_i, and the solutions are given in the first part of Theorems 5(M) and 5(O). However, with power RFFs they are not solvable. For example, with $n = 2$ the power function version of Equation A13 is

$$b_1^{(m)} + (d_2/d_1)^{1/a_2-1}(b_1^{(m)})^{a_1-1/a_2-1} = k$$

and clearly not solvable in elementary functions for $b_1^{(m)}$.

Once the response functions $f_i^{(m)}$, $f_i^{(o)}$ are found by solving Equations A13 and A14, the demand functions in the second part of Theorems 5(M) and 5(O) follow by substituting the response functions into the RFF:

$$r_i^{(m)} = h_i(b_i^{(m)}) = h_i(f_i^{(m)}),$$

$$r_i^{(o)} = h_i(b_i^{(o)}) = h_i(f_i^{(o)}).$$

Theorem 2

For hyperbolic RFFs, it is clear that $h_i'(b_i) = a_i(r_i/b_i)^2$; therefore, the pairwise condition for maximization, $h_i'(b_i^{(o)}) = h_j'(b_j^{(o)})$, directly implies Theorem 2.

With power functions we get essentially the same relationship (Rachlin, 1978). The marginal RFF is equal to $a_i r_i/b_i$, and the pairwise condition implies $b_i^{(o)}/b_j^{(o)} = (a_i/a_j)r_i^{(o)}/r_j^{(o)}$.

Theorem 6

Theorem 6 follows from Equation 34 and Assumption 4.

Theorem 7

To prove Theorem 7 notice that Assumptions 1, 2, and 3 imply (after rescaling) that $0 < b_i < k$ for all i. Rewriting the response functions slightly, one obtains the following two pairs of inequalities:

$$0 < \frac{d_i^{(m)}k + a_i^{(m)} - a_i}{d_i^{(m)} + d_i} < k,$$

$$0 < \frac{d_i^{(o)}k + a_i^{(o)} - (a_i^{(o)}a_i)^{1/2}}{d_i^{(o)} + (a_i^{(o)}a_i)^{1/2}d_i} < k.$$

These can be rewritten as

$$kd_i < (a_i - a_i^{(m)}) < kd_i^{(m)},$$

$$kd_i a_i^{-1/2} < [a_i^{-1/2} - (a_i^{(o)})^{1/2}] < kd_i^{(o)}(a_i^{(o)})^{-1/2},$$

which lead directly to Theorem 7(M) and Theorem 7(O).

Theorem 4: Aggregation

In general, aggregation of the RFF with a matching/maximizing response rule will hold if for any aggregate response class Q,

$$\sum_{i\in Q} h_i[h_i^{*\,-1}(\lambda)] = h_Q[\sum_{i\in Q} h_i^{*\,-1}(\lambda)]$$

$$[\text{because } b_i^{(m)} = h_i^{*\,-1}(\lambda)], \qquad \text{(A15)}$$

$$\sum_{i\in Q} h_i[h_i'^{\,-1}(\lambda)] = h_Q[\sum_{i\in Q} h_i'^{\,-1}(\lambda)]$$

$$[\text{because } b_i^{(o)} = h_i'^{\,-1}(\lambda)], \qquad \text{(A16)}$$

and if the aggregate RFF h_Q belongs to the same class as the individual RFFs h_i. To obtain the condition necessary for aggregation of the marginal RFF, one only has to replace h_i and h_Q in Equations A15 and A16 with the marginal RFFs h'_i and h'_Q. Clearly, aggregation is a stronger property than marginal aggregation.

Theorem 4(M)

The proof is much simplified if the hyperbolic RFF is reparameterized as follows: Let $s_i = 1/d_i$ and $c_i = a_i/d_i$. With these new parameters the hyperbola is written as

$$r_i = \frac{s_i b_i}{b_i + c_i}. \tag{A17}$$

A bit of algebraic manipulation will show that the formulas defining aggregate parameters $d_Q^{(m)}$ and $a_Q^{(m)}$ in Theorem 4(M) imply simple summation of the new parameters s_i and c_i:

$$\begin{aligned} s_Q^{(m)} &= \sum_{i\in Q} s_i, \\ c_Q^{(m)} &= \sum_{i\in Q} c_i. \end{aligned} \tag{A18}$$

From Equation A5 each $b_i^{(m)}$ can be written as a function of λ:

$$b_i^{(m)}(\lambda) = h_i^{*-1}(\lambda) = (s_i/\lambda) - c_i. \tag{A19}$$

This permits one, in turn, to also write each $r_i^{(m)}$ as a function of λ:

$$r_i^{(m)}(\lambda) = h_i[b_i^{(m)}(\lambda)] = s_i - \lambda c_i. \tag{A20}$$

Summing up Equations A19 and A20 over all i in Q

$$\begin{aligned} b_Q^{(m)} = \sum_{i\in Q} b_i^{(m)} &= \sum_{i\in Q} s_i/\lambda - \sum_{i\in Q} c_i \\ &= (s_Q^{(m)}/\lambda) - c_Q^{(m)}, \end{aligned}$$

$$r_Q^{(m)} = \sum_{i\in Q} r_i^{(m)} = \sum_{i\in Q} s_i - \lambda \sum_{i\in Q} c_i = s_Q^{(m)} - \lambda c_Q^{(m)}.$$

By eliminating λ from this pair of equations one proves

$$r_Q^{(m)} = \frac{s_Q^{(m)} b_Q^{(m)}}{b_Q^{(m)} + c_Q^{(m)}} = \frac{b_Q^{(m)}}{d_Q^{(m)} b_Q^{(m)} + a_Q^{(m)}}.$$

Theorem 4(O)

We use the same trick of reparametrizing the hyperbola, this time with $c_i = a_i/d_i$ and $e_i = (\sqrt{a_i})/d_i$. Again, the equations defining $a_Q^{(o)}$ and $d_Q^{(o)}$ in Theorem 4(O) will hold if and only if

$$\begin{aligned} c_Q^{(o)} &= \sum_{i\in Q} c_i, \\ e_Q^{(o)} &= \sum_{i\in Q} e_i. \end{aligned} \tag{A21}$$

In new guise the hyperbolic RFF is written as

$$r_i = \frac{(e_i^2/c_i) b_i}{b_i + c_i}$$

and the marginal RFF as

$$\frac{dr_i}{db_i} = \left(\frac{e_i}{b_i + c_i}\right)^2. \tag{A22}$$

From Equation A6 one can write $b_i^{(o)}$ and $r_i^{(o)}$ as functions of λ:

$$b_i^{(o)}(\lambda) = (e_i/\sqrt{\lambda}) - c_i, \tag{A23}$$

$$r_i^{(o)}(\lambda) = (e_i^2/c_i) - e_i\sqrt{\lambda}. \tag{A24}$$

Summing Equations A23 and A24 over all i in Q,

$$b_Q^{(o)} = (e_Q^{(o)}/\sqrt{\lambda}) - c_Q^{(o)},$$

$$r_Q^{(o)} = \sum_{i\in Q} (e_i^2/c_i) - e_Q^{(o)}\sqrt{\lambda},$$

and eliminating $\sqrt{\lambda}$ from the pair,

$$r_Q^{(o)} = \frac{(b_Q^{(o)} + c_Q^{(o)}) \sum_{i\in Q} (e_i^2/c_i) - (e_Q^{(o)})^2}{b_Q^{(o)} + c_Q^{(o)}}. \tag{A25}$$

From this one sees, incidentally, that aggregation of the RFF fails under maximization because of the presence of an additive constant in the numerator. Differentiating Equation A25 with respect to $b_Q^{(o)}$ yields

$$\frac{dr_Q^{(o)}}{db_Q^{(o)}} = \left(\frac{e_Q^{(o)}}{b_Q^{(o)} + c_Q^{(o)}}\right)^2,$$

which by virtue of Equation A22 is sufficient to prove Theorem 4(O).

It is perhaps interesting to note that matching and pairwise linearity are, without any auxiliary assumptions about the shape of the RFFs, a sufficient condition for aggregation (Theorem 4[M]). A sketch of the proof would point out two key steps. The first one involves showing that if matching equilibria with some set of RFFs produces pairwise linearity, then it must be possible to find a function ϕ such that every RFF in the set can be written as

$$h_i(b_i) = b_i \phi(d_i b_i + a_i).$$

For hyperbolas this function ϕ is $\phi(x) = 1/x$. The reader can verify that ϕ cancels out in the derivation of the matching equilibria (Equations A7, A9, and A13), which also shows that the response function in Theorem 5(M)—with its parameters uninterpreted, however—is the response function for any RFF system that produces pairwise linearity.

The second step is to show that aggregation of the RFFs h_i follows from the aggregation of the inverse local RFFs h_i^{*-1}. In the present case, if $h_i(b_i) = b_i\phi(d_ib_i + a_i)$ or, reparametrizing, $h_i(b_i) = b_i\phi(b_i/s_i + c_i/s_i)$, then

$$b_i^{(\mathrm{m})} = h_i^{*\,-1}(\lambda) = s_i\phi^{-1}(\lambda) - c_i$$

and

$$b_Q^{(\mathrm{m})} = s_Q^{(\mathrm{m})}\phi^{-1}(\lambda) - c_Q^{(\mathrm{m})}$$

or

$$\lambda = \phi(b_Q^{(\mathrm{m})}/s_Q^{(\mathrm{m})} + c_Q^{(\mathrm{m})}/s_Q^{(\mathrm{m})}).$$

Now, essentially repeating the proof of Theorem 4(M) after Equation A19

$$\begin{aligned}\sum_{i\in Q} r_i^{(\mathrm{m})} &= \sum_{i\in Q} h_i[h_i^{*\,-1}(\lambda)]\\ &= \sum_{i\in Q} b_i^{(\mathrm{m})}h_i^*[h_i^{*\,-1}(\lambda)]\\ &= \lambda b_Q^{(\mathrm{m})}\\ &= b_Q^{(\mathrm{m})}\phi[(b_Q^{(\mathrm{m})}/s_Q^{(\mathrm{m})}) + (c_Q^{(\mathrm{m})}/s_Q^{(\mathrm{m})})].\end{aligned}$$

Is pairwise linearity also a necessary condition for aggregation? No, aggregation is a much weaker property of RFF systems. To identify a plausible necessary condition for aggregation one has to specify more precisely what is meant by the aggregate RFF h_Q being of the "same class" as the individual RFFs h_i. Let us be strict and require that if each h_i has ℓ parameters $p_i^1, \ldots, p_i^\ell$, so that

$$h_i(b_i) = h(b_i; p_i^1, \ldots, p_i^\ell),$$

then the aggregate RFF is

$$r_Q = h(\sum_{i\in Q} b_i; \sum_{i\in Q} p_i^1, \ldots, \sum_{i\in Q} p_i^\ell).$$

As before, the key step is the aggregation of the inverse local RFF:

$$\sum_{i\in Q} h^{*\,-1}(\lambda; p_i^1, \ldots, p_i^\ell) \overset{?}{=} h^{*\,-1}(\lambda; \sum_{i\in Q} p_i^1, \ldots, \sum_{i\in Q} p_i^\ell).$$

For this to be true, one must be able to write each $h_i^{*\,-1}$ as

$$\begin{aligned}h_i^{*\,-1}(\lambda) &= h^{*\,-1}(\lambda; p_i^1, \ldots, p_i^\ell)\\ &= \sum_{j=1}^{\ell} p_i^j\psi^j(\lambda),\end{aligned}$$

where the ψ^j are unspecified functions of λ.

This characterization of aggregation also provides a simple test of whether aggregation will hold for a particular family of RFFs. For the power function aggregation does not hold, because the inverse local RFF does not have the required form

$$h_i^{*\,-1}(\lambda) = (\lambda/d_i)^{1/(a_i-1)}$$

(unless all the a_i are identical).

Appendix B

Generalization of Theorem 3

This appendix contains a proof of the general version of Theorem 3. Briefly, this theorem states that for any class of "acceptable" feedback functions, the three response rules—matching, maximizing, and constant ratio (choice axiom)—either predict three entirely distinct patterns of responding or, as the other possibility, predict identical behavior.

We begin by defining an acceptable RFF.

DEFINITION. An RFF h_i is *acceptable* if it satisfies the following four conditions:

(i) *h_i is continuous, and has continuous first and second derivatives for all $x \geq 0$;*

(ii) *h_i is increasing: $h_i'(x) > 0$, for all $x \geq 0$;*

(iii) *h_i is marginally decreasing: $h_i''(x) < 0$, for all $x \geq 0$;*

(iv) *h_i offers no free reinforcement:* $h_i(0) = 0$.

Only Conditions **(iii)** and **(iv)** require comment. The strict inequality in Condition **(iii)** is not really necessary, but it makes the proof simpler by eliminating certain problems associated with linear (ratio schedule) RFFs. For example, both matching and maximizing solutions are indeterminate when two or more identical ratio schedules are present in an experiment. It is more convenient, therefore, to state how the theorem applies to experiments containing linear RFFs at the very end, instead of carrying them along as a troublesome special case.

Condition **(iv)**, on the other hand, is really needed for the proof to go through. For example, if the RFFs are logarithmic $h_i(b_i) = c_i \log b_i$, then the maximizing subject will follow the constant-ratio rule (i.e., $b_i^{(o)}/b_j^{(o)} = c_i/c_j$), but the matching subject will not.

From (i)–(iv) one can derive four elementary properties:

(v) *The marginal and local rates of reinforcement become equal as response rate goes to zero*: $h_i^*(0) = h_i'(0)$. Proof:

$$h_i'(0) = \lim_{x \to 0} \frac{h_i(x) - h_i(0)}{x} = \lim_{x \to 0} \frac{h_i(x)}{x} = h_i^*(0).$$

[By (i), h_i' is continuous at zero.]

(vi) *For any response rate $x > 0$, the local reinforcement rate is greater than the marginal reinforcement rate*: $h_i^*(x) > h_i'(x)$. Proof: By definition of h_i^*,

$$x h_i^*(x) = h_i(x) = \int_0^x h_i'(y)dy.$$

But $h_i''(y) < 0$ then implies $h_i^*(x) > h_i'(x)$.

(vii) *An alternative expression for h_i' is*

$$h_i'(x) = h_i^*(x) + x h_i^{*\prime}(x),$$

where $h_i^{*\prime}$ is the derivative of h_i^*. Proof:

$$h_i'(x) = \frac{d}{dx}[x h_i^*(x)].$$

(viii) *The local rate of reinforcement is a decreasing function of response rate.* Proof: From (vii) it follows that

$$h_i^{*\prime}(x) = \frac{1}{x}[h_i'(x) - h_i^*(x)] < 0,$$

because $h_i'(x) < h_i^*(x)$ by (vi) above.

An experiment (E) is fully specified by a list (L) of schedules in that experiment and by a value (k) for the total amount of behavior that the subject can put out.

A well-defined response rule, then, assigns to each experiment E = (L, k) a list of predicted response rates: $\{b_{i,\mathrm{E}};\ i \in L\}$.

Let Γ be a class of acceptable RFFs. First, a response rule is consistent with the choice axiom for class Γ if and only if for any two RFFs i, j in Γ one of the following two conditions holds:

1. Absolute preference: In any experiment E containing RFFs i and j

$$b_{j,\mathrm{E}} = 0$$

or, in any experiment E containing RFFs i and j,

$$b_{i,\mathrm{E}} = 0.$$

2. Constant-ratio rule: There exists a constant $\omega > 0$ such that in any experiment E containing RFFs i and j either

$$\frac{b_{i,\mathrm{E}}}{b_{j,\mathrm{E}}} = \omega$$

or

$$b_{i,\mathrm{E}} = b_{j,\mathrm{E}} = 0.$$

Second, matching and maximizing solutions are identical for a class of RFFs Γ if and only if for any two schedules i, j in Γ and any experiment E containing i and j, the matching equilibria, $b_{i,\mathrm{E}}^*$ and $b_{j,\mathrm{E}}^*$, maximize the rate of reinforcement obtained from schedules i and j subject to the constraint $b_i + b_j = b_{i,\mathrm{E}}^* + b_{j,\mathrm{E}}^*$. Where necessary I will supplement matching with the melioration condition (Appendix C) so that there is no ambiguity about the direction of exclusive preference.

The theorem we wish to prove then asserts the following:

THEOREM. *For any class of acceptable schedules Γ, conditions (A), (B), and (C) below are equivalent:*

(A) *Matching is consistent with the choice axiom for class Γ.*

(B) *Maximizing is consistent with the choice axiom for class Γ.*

(C) *Matching and maximizing solutions are identical for class Γ.*

The mathematical content of this theorem is exhausted by two propositions, which will be stated now and derived in the course of the proof. The first one is used in proving the theorem for the case of absolute preference, and the second one is needed for the constant-ratio rule.

PROPOSITION 1. *A function f dominates a function g ($f > g$) if for all $x, y \geq 0$, $f(x) > g(y)$. Then for any h_i, h_j: $h_i^* > h_j^*$ if and only if $h_i' > h_j'$.*

PROPOSITION 2. *Conditions (a), (b), and (c) are equivalent:*

(a) *There exists $\omega > 0$ such that $h_i^*(x) = h_j^*(\omega x)$.*

(b) *There exists $\omega > 0$ such that $h_i'(x) = h_j'(\omega x)$.*

(c) *There exists a continuous function $\xi(x)$ such that $h_i^*(x) = h_j^*[\xi(x)]$ and $h_i'(x) = h_j'[\xi(x)]$.*

The proof of the theorem is broken into three sections.

PROOF. (A) implies (B) and (C).

First, consider the case where matching implies pairwise absolute preference. This will only be true when one local RFF dominates another: $h_i^* > h_j^*$. [If, on the contrary, $h_i^*(x) = h_j^*(y)$ for some x, y, then the experiment E = (L, k) such that L = $\{i, j\}$ and $k = x + y$ would produce a violation of absolute preference. Assertions like this will be made without proof later on.] By proving that $h_i' > h_j'$, one can show that maximization will also imply absolute preference and, furthermore, that the matching and maximizing solutions are for this case identical. Pick a constant (function) σ such that $h_i^* \geqslant \sigma \geqslant h_j^*$. From (**vi**) it follows that $\sigma > h_j'$. Suppose now that for some x, $h_i'(x) < \sigma$. But for any $y > x$, the fact that $h_i'' < 0$ implies

$$h_i^*(y) < \frac{1}{y}\left[\int_0^x h_i'(z) + \sigma(y - x)\right],$$

which, in the limit as $y \rightarrow \infty$, produces a contradiction, $h_i^*(y) < \sigma$. So, the marginal RFF h_i' must dominate h_j', as required.

If matching implies the constant-ratio rule for RFFs i and j, then the local rates of reinforcement for any response rate pair $b_i = x$, $b_j = \omega x$ have to be equal:

$$h_i^*(x) = h_j^*(\omega x). \tag{B1}$$

Taking derivatives with respect to x:

$$h_i^{*\prime}(x) = \omega h_j^{*\prime}(\omega x). \tag{B2}$$

Thus,

$$\begin{aligned} h_i'(x) &= h_i^*(x) + xh_i^{*\prime}(x) && [\text{by (vii)}] \\ &= h_j^*(\omega x) + x\omega h_j^{*\prime}(\omega x) && [\text{by Equations B1 and B2}] \\ &= h_j'(\omega x) && [\text{by (vii)}]. \end{aligned}$$

Therefore, the marginal rates of reinforcement at $b_i = x$, $b_j = \omega x$ are equal, maximization implies the constant-ratio rule, and the matching and maximizing solutions are identical.

Proof. (B) implies (A).

Taking the case of absolute preference first, all that is necessary is to show that dominance of marginal RFFs implies dominance of local RFFs. Assume, therefore, that $h_i' > h_j'$ and select σ such that $h_i' \geqslant \sigma \geqslant h_j'$. Because $h_j'(0) = h_j^*(0)$ by (**v**), and $h_j^*(x) \leq h_j^*(0)$ by (**viii**), $\sigma \geq h_j^*$. Now if for some x, $h_i^*(x) \leq \sigma$, then, by (**vi**), $h_i'(x) < h_i^*(x) \leq \sigma$, which is a contradiction. So, $h_i^* > \sigma \geq h_j^*$.

If, however, maximizing implies the constant-ratio rule, then the marginal rates of reinforcement for $b_i = x$, $b_j = \omega x$ have to be equal: $h_i'(x) = h_j'(\omega x)$. Because $h_j(0) = 0$ by (**iv**), one can write

$$h_j^*(x) = \frac{1}{x}\int_0^x h_j'(y)dy. \tag{B3}$$

By assumption, $h_j'(y) = h_i'(y/\omega)$. A change of variables, $z = y/\omega$, makes it possible to write Equation B3 as

$$\frac{\omega}{x}\int_0^{x/\omega} h_i'(z)dz = h_i^*(x/\omega),$$

which proves that matching behavior conforms to the constant-ratio rule: $h_i^*(x) = h_j^*(\omega x)$.

Proof. (C) implies (A).

First, consider the case where there exists some experiment in which the matching and maximizing solutions for responding on schedules i, j are positive. The marginal and local rates of reinforcement evaluated at the equilibrium response-rate pair, $b_i = x$ and $b_j = y$, have to be equal: $h_i^*(x) = h_j^*(y)$, $h_i'(x) = h_j'(y)$. Because the marginal RFFs h_i', h_j' are continuous [by (**i**)] and invertible [by (**iii**)], y can be written as a locally continuous function of x:

$$y = h_j'^{-1}[h_i'(x)] = \xi(x),$$

and one can reformulate the necessary matching and maximizing equilibrium conditions as

$$h_i^*(x) = h_j^*[\xi(x)], \tag{B4}$$

$$h_i'(x) = h_j'[\xi(x)]. \tag{B5}$$

The local continuity of ξ permits differentiation of Equation B4:

$$h_i^{*\prime}(x) = \xi'(x)h_j^{*\prime}[\xi(x)]. \tag{B6}$$

By (**vii**), the equality (Eq. B5) can be written as

$$h_i^*(x) + xh_i^{*\prime}(x) = h_j^*[\xi(x)] + \xi(x)h_j^{*\prime}[\xi(x)],$$

which, taking advantage of Equations B4 and B6, reduces to

$$x\xi'(x) = \xi(x).$$

Therefore, $\xi(x) = \omega x$, for some $\omega > 0$, and so both matching and maximizing will imply the constant-ratio rule for this pair of schedules.

The only other way that matching and maximizing solutions could agree would be for both to predict either b_i or b_j equal to zero in all experiments. In that case, however, both response rules would satisfy the absolute preference component of the choice axiom.

This completes the proof of the theorem.

Characterization of RFF Families for Which (A), (B), and (C) Hold

The constant-ratio rule is clearly a transitive relation over pairs of schedules: If relative response rate for RFFs i, j is constant in all experiments, and if the same is true for RFFs j, k, then relative response rate for i, k will also be constant in all experiments. This suggests that any class Γ for which (A), (B), and (C) hold may be broken up into disjoint subsets Γ^1, Γ^2, . . . such that for any schedule pair within one such subset matching (or maximizing) will predict the constant-ratio rule; whereas for any schedule pair taken from two distinct subsets matching will predict absolute preference. Furthermore, the local RFFs for schedules within one subset (J) can be written in terms of some function (F) specific to that subset:

$$h_i^*(b_i) = F_J(b_i/v_i), \quad v_i > 0. \tag{B7}$$

This is merely a generalization of Condition (a) in Proposition 2 to all schedules in Γ^J. The constants v_i, by the way, are the v-scale values that one would assign to individual schedules in Γ^J if one was applying the choice axiom directly to performance on schedules of reinforcement (Luce, 1959).

If Equation B7 is true, then the feedback functions are of the form

$$h_i(b_i) = b_i F_J(b_i/v_i). \tag{B8}$$

The relation between local RFFs taken from different subsets of Γ must of course be one of dominance.

One can now see how ratio schedules can be fitted within this representational scheme. The class Γ may contain any number of distinct ratio schedules provided that the local RFF for any particular ratio schedule (which is a constant function) either dominates or is dominated by any other local RFF taken from Γ. In other words, ratio schedules can be a part of a class satisfying (A), (B), and (C) only if they stand in the relation of absolute preference to all other schedules.

Appendix C

Meliorating Equilibria

This appendix deals with one remaining lacuna in the matching model (Assumptions 1-5[M], plus rescaling), namely, What predictions will it make when the constraints on the parameter space as defined in Theorem 5 are violated and a matching equilibrium satisfying Assumptions 1-5(M) does not exist? The concurrent VR-VR experiment will serve as the prototypical example of this problem.

From a purely formal standpoint, exclusive preference for either alternative would satisfy the definition of matching as it is given in Assumption 5(M). However, the evidence (Herrnstein & Loveland, 1975; Vaughan, 1976) suggests that subjects always go to the higher valued ratio schedule. Herrnstein and Vaughan (1980) have proposed that what the subjects actually do in all concurrent-schedule experiments is shift their responding in a gradual fashion from alternatives having low local rates of reinforcement to alternatives with high local rates. This process, called *melioration*, would reach equilibrium at the usual matching response distributions. Moreover, it would rule out the never-observed matching equilibria, such as exclusive preference for the lower valued ratio schedule.

If we accept melioration as the proper extension of matching to situations where exclusive preference is unavoidable, the problem here is to formulate a set of conditions that would define the meliorating equilibria in the same way that Assumptions 1-5(M) defined the matching ones. I propose the following set of inequalities as the defining ones (they are somewhat related to the Kuhn-Tucker conditions in nonlinear programming; Intriligator, 1971):

$$\begin{aligned} &\sum_{i=1}^{n} b_i = k; \\ &b_i \geq 0, \quad i = 1, \ldots, n; \\ &\frac{r_i}{b_i} \leq \lambda, \quad i = 1, \ldots, n; \\ &\sum_{i=1}^{n} b_i\left(\frac{r_i}{b_i} - \lambda\right) = 0. \end{aligned} \tag{C1}$$

There are two innovations: (a) The local rates r_i/b_i may be less than or equal to some constant λ; (b) but, for those schedules for which the local rate is actually less than λ, the response rate b_i has to equal zero or otherwise the last equation would not hold.

Applying the system (Eq. C1) to the concurrent VR-VR example, with $r_1 = b_1/a_1$, $r_2 = b_2/a_2$, and $a_2 > a_1$, one obtains a unique solution: $b_1 = k$ and $b_2 = 0$ (and $\lambda = 1/a_1$), as required.

Received July 17, 1981
Revision received January 19, 1982 ■