JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1988, 49, 303–320 NUMBER 2 (MARCH)

QUASI-DYNAMIC CHOICE MODELS: MELIORATION AND RATIO INVARIANCE

J. E. R. STADDON

RUHR-UNIVERSITÄT AND DUKE UNIVERSITY

There is continuing controversy about the behavioral process or processes that underlie the major regularities of free-operant choice such as molar matching and systematic deviations therefrom. A recent interchange between Vaughan and Silberberg and Ziriax concerned the relative merits of melioration, and a computer simulation of molecular maximizing. There are difficulties in evaluating theories expressed as computer programs because many arbitrary decisions must often be made in order to get the programs to operate. I therefore propose an alternative form of model that I term *quasi-dynamic* as a useful intermediate form of theory appropriate to our current state of knowledge about free-operant choice. Quasi-dynamic models resemble the game-theoretic analyses now commonplace in biology in that they can predict stable and unstable equilibria but not dynamic properties such as learning curves. It is possible to interpret melioration as a quasi-dynamic model. An alternative quasi-dynamic model for probabilistic choice, *ratio invariance*, has been proposed by Horner and Staddon. The present paper compares the predictions of melioration and ratio invariance for five experimental situations: concurrent variable-interval variable-interval schedules, concurrent variable-interval variable-ratio schedules, the two-armed bandit (concurrent random-ratio schedules), and two types of frequency-dependent schedule. Neither approach easily explains all the data, but ratio invariance seems to provide a better picture of pigeons' response to probabilistic choice procedures. Ratio invariance is also more adaptive (less susceptible to "traps") and closer to the original expression of the law of effect than pure hill-climbing processes such as momentary maximizing and melioration, although such processes may come in to play on more complex procedures that provide opportunities for temporal discrimination.

Key words: Bush-Mosteller, melioration, momentary maximizing, molecular maximizing, ratio invariance, source independence, matching

---

The initial stimulus for this paper was Stephen Lea's request to the reviewers of a manuscript by Will Vaughan that they prepare commentaries suitable for independent publication along with Vaughan's paper, which is itself a critical commentary on an earlier paper by Silberberg and Ziriax (1985). The most important aspect of this controversy is about the adequacy of different models for free-operant choice, which is the major focus of this paper.

To decide between competing choice theories it is essential to frame them in a quantitataive way that permits clear and definite predictions in complex situations. A common way to do this is computer simulation, the method chosen by Silberberg and Ziriax (1985). But the simulation route, though preferable to imprecise verbal statements, has its drawbacks. For example, it forces one to make specific assumptions (e.g., about the temporal distribution of responses or the size of short-term memory) that are essential if the simulation is to operate, but may be irrelevant to the critical differences between theories. If one theory fails and the other succeeds, it may be because one is true and the other false—or because of some unsuspected difference in the details of the simulations. When only one of the two theories is in the form of a simulation, as in Silberberg and Ziriax's comparison between their simulated theory and Vaughan's (1981) verbally stated one, these problems are enhanced. There are also general objections to computer programs as a form of theory, having to do with their complexity (see Richerson & Boyd, 1987, for an excellent discussion of complexity vs. simplicity in theory). Some quantitative form that is both simpler and more general than a computer program is preferable.

Vaughan (1985) has provided the basis for

I thank Juan Delius and the three reviewers (Stephen Lea, Tony Nevin, and Will Vaughan) for invaluable comments on the manuscript. I am especially grateful to Joachim Krauth, who saved me from many mathematical errors and infelicities—I take full responsibility for those that remain. Research was supported by grants to Duke University from the National Science Foundation and the Pew Memorial Trust and to J. D. Delius by the Deutsche Forschungsgemeinschaft. I thank the Alexander von Humboldt-Stiftung for support during the sabbatical leave when this work was done. Requests for reprints should be sent to the author, Department of Psychology, Duke University, Durham, North Carolina 27706.

a formal analysis of melioration. With his simplifying assumptions it is possible to write down a formal model of melioration of a type somewhere between the simple static models traditional in operant psychology and the full-blown dynamic models of physics. It seems useful to have a name for models of this sort, so I call them *quasi-dynamic*. A second model of this type is one that John Horner and I have worked on for the past 3 years that we term *ratio invariance* (Horner & Staddon, 1985, 1987). Ratio invariance can usefully be compared with melioration. Molecular maximizing, the computer-simulation model proposed by Silberberg and Ziriax, does not lend itself as easily to analysis in this way, however. This paper therefore discusses in detail only melioration and ratio invariance.

The paper is presented in four parts. The first part is a brief general discussion of static, dynamic, and quasi-dynamic choice theories. The second part is a formal analysis of melioration and ratio invariance in which predictions are derived for five types of experimental situations (concurrent variable-interval variable-interval schedules, concurrent variable-interval variable-ratio schedules, the two-armed bandit, the frequency-dependent procedure studied by Vaughan and Silberberg and Ziriax, and a frequency-dependent ratio procedure studied by Horner and Staddon). The third section compares these predictions with standard experimental results from studies with hungry pigeons and food reinforcement. In the course of this comparison, I comment on some of the methodological issues raised by the controversy beween Vaughan and Silberberg and Ziriax. The final section summarizes the relative merits of melioration and ratio invariance.

Types of Choice Theory

Theories of behavior are of two main types: *end-state* (*static*) theories and *dynamic* theories. Examples of static theories of operant choice behavior are the matching law (Herrnstein, 1961) and the various molar optimality models such as minimum-distance (Staddon, 1979) or economic accounts (e.g., Lea, 1981; Rachlin, 1978; Rachlin, Green, Kagel, & Battalio, 1976). Examples of dynamic theories are the kinetic theory (Myerson & Miezin, 1980) and learning theories derived from the classical Bush and Mosteller (1955) linear-operator model (e.g., Sternberg, 1963; Vaughan, 1982). Melioration (Herrnstein & Vaughan, 1980) is a partially specified dynamic model.

End-state theories merely specify some property of the behavioral steady state, such as matching of response and reinforcement ratios, or optimization of an objective function (e.g., maximization of reinforcement rate). These theories make no specification of the process by which this end state is to be achieved. Dynamic theories, on the other hand, specify the underlying process, either moment-by-moment, trial-by-trial, or response-by-response. Dynamic theories predict both the possible end states and the paths by which the end states are reached.

There is another type that appears not to fit into either of these two categories, namely *decision-rule* theories. Examples of decision-rule theories are momentary maximizing (Shimp, 1966; Staddon, Hinson, & Kram, 1981) and molecular maximization (Silberberg & Ziriax, 1985). These theories do make predictions on a response-by-response basis: momentary maximizing, for example, asserts that when a response is made, it will be to the alternative offering the highest reward probability. (It is not correct to define momentary maximizing, as Vaughan (1987) does, as ". . . the theory that a changeover occurs from one side to the other when the latter has a higher probability of reinforcement than the former . . ." (p. 333), because this condition normally obtains right after *any* response, so that momentary maximizing defined in this way would always predict alternation at the highest possible response rate.) But decision-rule theories are not true dynamic theories, because they do not specify whether a response will or will not occur. To simulate decision-rule theories, *some* assumption must be made about the temporal distribution of responses—but such an assumption is necessarily arbitrary and not integral to the theory. Rather, these theories specify a property that steady-state responding must have. In this sense they are closer to static theories than dynamic ones.

Dynamic processes can show a variety of steady-state behavior patterns (the summary here is much simplified; see Crutchfield, Farmer, Packard, & Shaw, 1986, or Weisbuch, 1986, for recent reviews of dynamic systems and related topics). (a) The simplest is *equilibrium behavior*. For example, if water is

poured steadily into a leaky bucket, the water level eventually stabilizes at a point determined by the rate of inflow and the size of the hole. This is an example of a *stable* equilibrium, because the effect of a small perturbation (such as taking some water out of the bucket) is eventually corrected. Equilibria can also be *unstable* (the usual example is a ball balanced on an upturned hemisphere): such equilibria are almost never observed, of course, because even a slight perturbation causes a self-reinforcing movement away from the equilibrium point. Occasionally, an equilibrium is *neutral,* that is, any perturbation has an effect that is retained (the usual example is a ball resting on a horizontal plane). (b) Many dynamic systems also show *periodic behavior* (i.e., simple or complex oscillation) that may eventually die out (think of a ball dropped into a hemispherical bowl, for example) or may persist indefinitely. (c) Perhaps the most interesting behavior of dynamic systems (one studied only relatively recently) is *chaos,* that is, apparently irregular, aperiodic behavior that nevertheless has a completely deterministic basis (e.g., Crutchfield et al., 1986; May, 1976).

I will argue that our understanding of free-operant choice is at present limited to models that may be termed *quasi-dynamic*: they allow us to say something about equilibrium behavior—whether there are equilibria or not, and whether or not they are stable—but lack features necessary for a full dynamic analysis. Models of this sort are not found in physics, where true dynamic analyses are nearly always possible, and as far as I know they have not been specifically discussed in psychology. The distinctive feature of such models is that they are not specified well enough to permit predictions about trajectories or periodic behavior, that is, about learning curves, sequential statistics, or responses to rapidly changing conditions. (It is perhaps worth remembering that predictions of learning and extinction curves preoccupied early mathematical learning theorists. The difficulty of predicting highly variable data of this type may account for the decline of interest in that approach.) Nevertheless, I hope to show that despite their limiations, quasi-dynamic models can yield predictions that are surprisingly detailed and permit powerful experimental tests.

Quasi-dynamic models, as I discuss them here, are Markovian, that is, they define the expected change in the dependent variable (usually, but not necessarily, response probability) as a function solely of its current level; there is no dependence on earlier values. In common with the great majority of learning models, they are first-order only.

Quasi-dynamic analysis is closely related to the game-theoretic modeling now commonplace in biology (see, e.g., Maynard Smith, 1982), in which the objective is to discover *evolutionarily stable strategies* (ESS). An ESS is an equilibrium characterized by local stability, that is, immunity to small perturbations ("An ESS is a strategy such that . . . no mutant strategy could invade the population under the influence of natural selection": Maynard Smith, 1982, p. 10). For example, if the ESS is a mixed strategy consisting of proportion s of Strategy A and $1 - s$ of Strategy B, then a small increase in the frequency of A (say) creates a restoring increase in the fitness of individuals following Strategy B. ESS theory is concerned with finding equilibria; it is not directly concerned with the dynamic behavior of the system (e.g., the details of the changes in gene frequencies from one generation to the next). ESS analysis arises naturally once we consider patterns of animal behavior as the outcome of constrained competition among behavioral strategies. In the same way, quasi-dynamic analyses of individual behavior are best tested by experimental situations in which the schedule has game-like properties.

The frequency-dependent schedules recently studied by Herrnstein and Vaughan (1980) and Horner and Staddon (1985, 1987), and discussed in more detail below, closely resemble what Schelling (1978) has termed the *multiperson prisoner's dilemma* (MPD), and are formally identical to the population biologist's *frequency-dependent selection.* Schelling (1978, p. 218) defines the MPD in terms of four properties: (a) There are N people (here, *responses*) each of whom has the same two choices and payoffs (rewards). (b) Each person (response) has a preferred choice, which is the same for all. This is an arbitrary feature, which is that at any choice proportion one side is always paid off at a higher rate than the other. Translated to the schedule situation it is equivalent to saying that at any level of preference, one side is always associated with a higher reward probability than the other. Horner and I have studied nontrivial frequency-dependent

schedules in which both alternatives are always paid off at the same reward probability and others in which the probability difference reverses from one choice proportion to another, so I see no need for this constraint. (c) Benefit to each individual increases the more individuals choose the unpreferred alternative. In other words, on frequency-dependent schedules the reward probability for a given choice is directly related to the frequency of the *other* choice. This is also an arbitrary constraint: the essence of these procedures is just *some* kind of dependence of the payoff for one or both choices on the relative frequency of each choice. The fourth defining property concerns coalition formation as it is determined by the arbitrary feature (b), above. It is not directly relevant to the single-individual frequency-dependent schedule, both because (b) need not hold there and because we assume that the animal's choice set is just the two response alternatives—it cannot directly select a given choice *proportion.*

The analogy with frequency-dependent selection is even more direct. The simplest example is the oldest, selection for the 50:50 sex ratio, which Fisher (1930) years ago explained by the increased fitness of individuals whose progeny were of the minority phenotype. This is a frequency-dependent schedule in which the "probability of reward" for each phenotype is inversely related to its frequency, with the curves crossing at the 50:50 point.

Stable points (i.e., steady-state sex ratio, or proportion of individuals choosing each alternative) for frequency-dependent selection or MPD situations are derived using the kind of quasi-dynamic analyses I describe here.

Melioration

Melioration was originally described in a purely verbal way (Herrnstein & Vaughan, 1980; see Rachlin, 1973, for a very similar idea applied to successive discrimination), but recently has received a more formal development at the hands of Vaughan (1985). The present analysis uses some of the same simplifications as Vaughan used. The theory of melioration has two parts, a causal variable and a behavior-change rule. The causal variable is *local reinforcement rate,* that is, the rate of reinforcement (reward) obtained during the time when a subject is actually responding to a given alternative. There are some uncertainties about precisely how to measure this period, but for present purposes it is sufficiently defined by the constraint that the more time the animal spends on one alternative, the less time will be available for the other. The behavior-change rule is that "If the local [reinforcement] rate on one side is higher than on the other . . . more time will be spent on the better side" (Vaughan, 1985, p. 385). This is a principle that resembles gas pressure (in which pressures in deformable compartments tend to equalize; see Staddon, 1982, for formal exploration of this resemblance) or heat transfer (in which heat flows from higher temperature to low).

It is possible to construe melioration in a way that allows us to predict much while assuming little. Consider the standard two-choice free-operant situation, with variable-(random) interval (VI) schedules associated with each choice. If the values of the VI schedules are chosen so that total reward rate is approximately constant (see Herrnstein, 1961), then total response rate is approximately constant (this is true over quite a wide range even if total reward rate varies; cf. Catania & Reynolds, 1968). Measurements have also shown that local response rate is approximately constant under these conditions (see discussion in Heyman, 1979, and Staddon & Motheral, 1979), so that the relevant dependent variable is the amount of time the animal spends responding to each alternative. For simplicity I assume a session of unit length, so that the times spent working on each side are t (on the right: R) and $1 - t$ (on the left: L). Local reward rates will then be functions of these local times: $r_R(t)$ (on the right) and $r_L(1 - t)$ (on the left). $r_R(t)$ and $r_L(1 - t)$ are feedback functions computed with respect to local rates; they are very different from the familiar molar feedback functions computed with respect to overall (i.e., whole-session) response and reward rates.

Notice that if t is very small, $r_R(t)$ can become very large; for example, if the animal makes only a single, rewarded, response on the right during an experimental session and spends the rest of his time responding on the left, then his local reward rate on the right will be extremely high—much greater than the VI rate (1/VI value).

Melioration says that delta(t), the expected tendency for choice proportion (relative time), t, to change, is some function, G, of the pre-

vailing difference between local reward rates (see Appendix for symbol definitions). Formally (Vaughan, 1985, Equation 8),

$$\text{delta}(t) = G[r_R(t) - r_L(1 - t)].$$

In other words, if local reward rate on the right is higher than local reward rate on the left, the animal spends a bit more time on the right.

To derive predictions from this principle we need to know more about function G, which tells us *how much* more time is allocated for a given difference in local reward rates. Two things about G are specified in Vaughan's verbal theory: (a) G is positive and monotonic; that is, the larger the difference between the local reward rates, the greater the tendency to allocate more time to the higher rate alternative, and (b) $G(0) = 0$ (i.e., G is zero when the rate difference is zero).

By themselves, these two assumptions do not seem to be sufficient to allow us either to solve or (more modestly) to find the equilibria for the basic melioration equation. It is possible to proceed by picking a particular class of functions for which (a) and (b) are true, however. I will therefore only consider the case in which G is some positive multiplier, which also provides a simple way to deal with the effects of *time window* (i.e., the period over which reward rates are computed by the animal). When the time window is short (local rates are computed by the animal only for the recent past), then any change in reward conditions has a rapid effect on behavior. This corresponds to a large value for G. When the time window is long, however, any change takes some time to have an effect. This corresponds to a small value for G. It turns out to be easy to arrive at equilibrium predictions that are invariant across any positive value for G. G may be any positive (not necessarily monotonic) function of t, for example. (Consider the general case in which $\text{delta}(t) = G(t)g(t, \mathbf{v})$, where G is a positive function of t alone and g is some other function of t and a vector of parameters, $\mathbf{v}$, such that when the difference between $r_R(t)$ and $r_L(1 - t)$ is zero, g is also zero (here $g = [r_R(t) - r_L(1 - t)]$). By the rule for products, the derivative of $\text{delta}(t)$ is therefore $d/dt[\text{delta}(t)] = G(t)d/dt[g(t, \mathbf{v})] + g(t, \mathbf{v})d/dt[G(t)]$. When g is zero, the second term vanishes, because $g(t, \mathbf{v})$ is equal to zero; hence, so long as $G(t)$ is positive, the sign of the derivative depends entirely upon the sign of the derivative of $g(t, \mathbf{v})$. Hence, the equilibrium properties of the process are independent of any positive multiplier.) Note that G must be *some* function of t, so that progressive increments or decrements in t do not drive t above one or below zero.

If we are interested only in equilibria, then two properties of $\text{delta}(t)$ are of interest: the t-values for which it is zero and the slope of the function at those points. Because for these purposes G can be any positive multiplier, we can without loss of generality let $G = 1$ in the subsequent analysis, so that the quasi-dynamic version of the melioration hypothesis amounts to

$$\text{delta}(t) = r_R(t) - r_L(1 - t), \qquad (1)$$

which is similar to the dynamic version of melioration proposed by Myerson and Hale (1984).

Concurrent variable-interval schedule. To derive predictions from this version of melioration it is necessary to define the (local) feedback function $r(t)$, which may not be a trivial matter because $r(t)$ will depend not just on t, but on how time is distributed across the experimental session. For example, on a concurrent VI VI schedule, if the animal spends only 3 s (say) out of an hour responding on the right, its value for $r(t)$ will be much higher if the 3 s occur as three 1-s blocks evenly spaced throughout the hour than if they occur as a single 3-s block. Despite this limitation, in common with other theorists I will assume that the temporal-distribution variable has similar effects on both choices and can therefore be neglected. The most salient property of $r(t)$ is that as t increases, $r(t)$ must decrease: as the animal allocates a larger fraction of the total time to one alternative, the more rewards it gets, but, because the schedule is time-based, the lower the (local) rate at which it gets them. But $r(t)$ will not normally decrease much below the value of the VI schedule (pigeons are quite efficient on VI). A simple possibility, therefore, is

$$r_R(t) = a/t, \qquad 0 \le t \le 1, \qquad (2)$$

where a is the scheduled rate (1/VI value) of the VI schedule on the right (this is the simplification adopted by Vaughan, 1985). In this situation, therefore, $r(t) = a$ when the animal spends the entire unit session responding to

the right choice. (Note that this feedback function assumes that animals respond fast enough, and allocate their responses between the two alternatives efficiently enough, that rewards are delayed a negligible amount of time after "setup." When this is the case, the *overall* (whole-session) reinforcement rate for each choice always equals the scheduled rate, as Equation 2 implies.) When $t < 1$, $r_R(t)$ will always be $>a$. Substituting Equation 2 in Equation 1 yields

$$\text{delta}(t) = a/t - b/(1 - t), \qquad (3)$$

where a and b are the scheduled VI rates for left and right choices, respectively. Equation 3 is the fundamental equation for melioration in the concurrent VI VI situation.

To find equilibria, we need to look at the values of t for which $\text{delta}(t) = 0$, if such values exist; and we need to know the slope of the $\text{delta}(t)$ function at those points. From Equation 3,

$$\text{delta}(t) \cdot t(1 - t) = a - t(a + b), \qquad (4)$$

so that when $\text{delta}(\hat{t}) = 0$, $\hat{t}$, the equilibrium value for t, is

$$\hat{t} = a/(a + b). \qquad (5)$$

From Equation 4 it is obvious that the slope of the function is negative at the equilibrium point, so that this is a stable equilibrium (because deviations from the equilibrium point cause opposing changes in $\text{delta}(t)$). Because t is proportion of total time, Equation 5 corresponds to perfect matching of time proportions to both obtained and scheduled reward proportions.

Concurrent variable-interval variable-ratio schedule. It is relatively easy to extend this analysis to concurrent VI VR schedules, with the right remaining a VI schedule, as before, but left now dispensing reward according to a random-ratio schedule with probability q (=1/ratio value). We assume the same basic $\text{delta}(t)$ relation as before (Equation 3), but substitute a different expression for the local reward rate on the ratio (left) side:

$$r_L(t) = qy_L,$$

where y_L is the local rate of ratio responding (assumed a constant) and q is the payoff probability. Substituting in Equation 1 then yields $\text{delta}(t) = a/t - qy_L$, which has a stable equilibrium at $\hat{t} = a/qy_L$. Given that the local response rate on the ratio alternative is y_L, then the obtained overall reward rate on that side, $R(y)$, is just $R(y) = qy_L(1 - \hat{t})$, so that $qy_L = R(y)/(1 - \hat{t})$, where y is overall response rate on the ratio side; substituting for qy_L in the equation for $\hat{t}$ and rearranging yields

$$\hat{t}/(1 - \hat{t}) = a/R(y) = R(x)/R(y), \qquad (6)$$

that is, simple matching of time-allocation ratios to obtained (overall) reward ratios.

Note that this prediction depends only on the assumption that the local response rate on the ratio side is constant, and not that it is the same as the local response rate on the variable-interval side. Because local response rate will usually be higher on the ratio than on the interval side (Herrnstein & Heyman, 1979), Equation 6 predicts biased matching in terms of response ratios, with more responses going to the ratio alternative. For example, suppose that the left-hand side of Equation 6 equals one ($t = 1 - t$), then in terms of relative responses the ratio will be <1, because local response rate on the ratio side is higher, so that Equation 6 implies a relation of the form

$$s/(1 - s) = kR(x)/R(y), \qquad 0 < k < 1,$$

where s is the proportion of choice responses on the interval alternative.

Simple as they are, these derivations may seem unnecessarily laborious, because it is obvious from Equation 1 that $\text{delta}(t) = 0$ when local reward rates are equal. We cannot know whether this equality is permitted (i.e., whether it is possible to obtain equal local reward rates on both sides), or, even if permitted, whether the equilibrium is stable, without knowing the feedback functions.

Two-armed bandit. To illustrate, consider as a third example the case of concurrent variable-(random)ratio schedules (the two-armed bandit problem). Here, local reward rates can be made equal only by responding more slowly on the higher probability (lower ratio) schedule, something not implied by any version of melioration. We can arrive at a more reasonable prediction by assuming that local response rates are equal, and the relevant dependent variable is then the proportion of responses, s, made to each alternative ($s = R/(R + L)$). If local response rate is equal to x on both sides, and the reward probabilities are p and q on R and L ($p > q$, by convention), then the local feedback functions are

$$r_R(s) = px, \text{ and } r_L(1 - s) = qx.$$

Substituting in Equation 1 yields

$$\text{delta}(s) = px - qx, \quad (7)$$

which is always positive. Hence, s increases until all responding is to the majority (here R) side, so that exclusive choice of the majority is the melioration prediction.

Asymmetrical frequency-dependent ratio schedules. As a fourth example, consider an asymmetrical frequency-dependent (AFD) random-ratio schedule studied by Horner and Staddon (1985, 1987). In this schedule, reward probability for each alternative depends on the animal's preference, S (where S = R/(R + L), as before). In these experiments, the value of choice proportion, S, is computed response by response in a moving "window" of M responses (see later discussion). The value of M, the averaging window, has some effects on the performance of the model (and the animal) but they are not relevant for the argument made here. For both alternatives the dependency is linear: on the right, for example, $p(S) = kS$, where k is a constant. But at any S-value, the reward probability on the left (the majority side) is always twice that on the right (the minority side) (i.e., $q(S) = 2p(S) = 2kS$). Under these conditions, the overall payoff probability is highest when $S = 1.0$ (exclusive choice of the minority side); that is, when in any run of M choices M are right and none are left. Overall payoff probability approaches zero as the animal tends toward exclusive choice of the left (majority) choice.

If S, the measured choice proportion, is an unbiased measure of s, choice probability, the melioration prediction here can be derived at once from Equation 7, by substituting S for s, kS for p, and $2kS$ for q: $\text{delta}(s) = kSx - 2kSx$, which is always negative. Hence melioration predicts fixation on the majority side ($S = 0$), which is highly suboptimal behavior.

Frequency-dependent interval schedule. As a final example, consider the frequency-dependent variable-interval schedule studied by Silberberg and Ziriax (1985) in their Condition 1. In this schedule the prevailing VI value on each side depends on the allocation of responding between sides. When responding is predominantly on the right, defined in a block-by-block window of either 6-s or 4-min duration (time proportion, $t >$ a critical value, t_{Rcrit}), there is a low VI schedule for R responses and a high one for L; when responding is predominantly on the left ($t < 1 - t_{Rcrit}$) there is a low VI schedule for L responses and a high one for R. Otherwise (in the middle region, when $t_{Rcrit} > t > 1 - t_{Rcrit}$), the same low VI operates for both choices. This procedure is illustrated in the top panel of Figure 1.

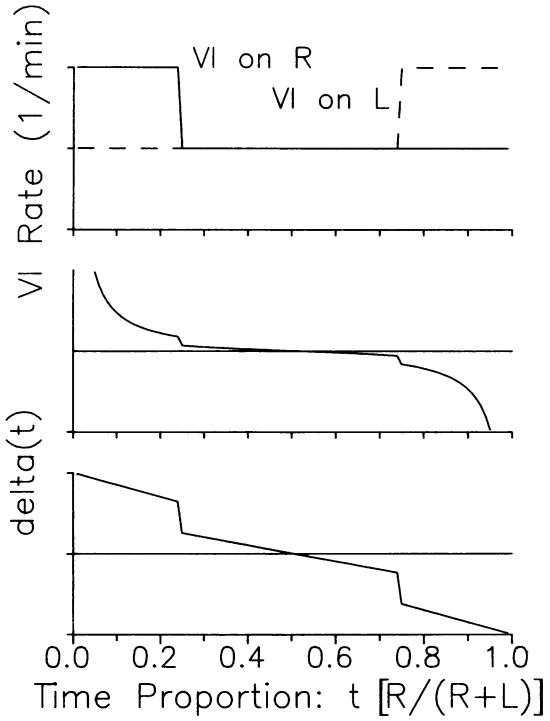

Fig. 1. Expected change in time allocation, delta(t), predicted by the quasi-dynamic version of melioration for the type of interdependent variable-interval schedule studied by Silberberg and Ziriax (1985, Condition 1): *high* VI value is twice the *low* value. Top panel: VI value as a function of time-allocation proportion, t (=R/(L + R)), for left and right responses. Middle panel: Three-part delta(t) function predicted by melioration (derived from Equation 4 in the text). Bottom panel: delta(t) function without $t(1 - t)$ multiplier (see text). Indifference (t = 0.5) is the stable point.

Formally, the feedback function for this schedule is defined by two values of the VI parameter:

$$r_R(t) = a/t, \qquad \text{for } t < t_{Rcrit}$$

and

$$r_R(t) = b/t \qquad \text{for } t > t_{Rcrit},$$

where a and b are maximum-scheduled re-

ward rates (i.e., 1/VI values), $a > b$, and $t_{\text{Rcrit}} = \frac{1}{4}$, $t_{\text{Lcrit}} = \frac{3}{4}$.

Because the VI schedule values in this procedure change depending on the prevailing preference (t-value), the delta(t) function derived from Equation 1 will have three different regions. For the middle region, in which both VI values are the same, from Equation 4

$$\text{delta}(t) \cdot t(1 - t) = b - t(b + b),$$

so that, ignoring the positive multiplier,

$$\text{delta}(t) = b(1 - 2t), \tag{8}$$

a stable equilibrium at $t = \frac{1}{2}$, indifference. For the region $0 < t < t_{\text{Rcrit}}$, $\text{delta}(t) \cdot t(1 - t) = a - t(a + b)$, which has an equilibrium at $\hat{t} = a/(a + b)$, as in Equation 5 above. This equilibrium is not realizable, however, because $a > b$, so that $\hat{t}$ is not within the region $0 < t < t_{\text{Rcrit}}$ over which this delta(t) function holds; this is similar for the symmetrical equilibrium at $\hat{t} = b/(a + b)$. The three-part delta(t) function associated with this procedure is illustrated in the bottom two panels in Figure 1. The middle panel shows the three parts, using Equation 4: $\text{delta}(t) = [a - t(a + b)]/t(1 - t)$; the bottom panel shows the function without the $t(1 - t)$ multiplier. Both graphs give the same information about equilibria. It is clear from Figure 1 that indifference, $t = \frac{1}{2}$, is the only stable point in this procedure, and is therefore the melioration prediction.

Ratio Invariance

John Horner and I have recently elaborated and tested experimentally a theoretical approach to probabilistic choice that we term ratio-invariant reward following (ratio invariance (RI); see Horner & Staddon, 1987, for the full analysis, which is merely summarized here). The approach is based upon relatively short-term experiments in a symmetrical two-choice situation (i.e., two physically identical responses, with a contingency that equates the effort of "staying" vs. "switching") reinforced according to similar, probabilistic schedules. We are fairly sure that RI is not the dominant process operating in well-trained animals in more complex situations—such as VI schedules—in which the existence of temporal discrimination, in the form of momentary maximizing or a related process, is well established. Nor will it apply (at least in the simple, symmetrical version discussed here) when there are more than two possible responses or when the two responses are not equivalent—in particular, when one "response" is a nonresponse, as in a single-choice situation (an asymmetrical version that might apply here is briefly described in connection with concurrent VI VR schedules). Nevertheless, because it is important to define the failures of a model as well as its successes, I discuss temporal as well as nontemporal situations.

Ratio invariance is an elementary stochastic implementation of the classical law of effect in a situation with two equivalent mutually exclusive and exhaustive response classes, with two constraints, source independence and effect-ratio invariance, that are explained below. The law of effect asserts that rewarded responses increase in probability and unrewarded responses decrease. Thus, the delta(s) (expected change in choice proportion) function has the general form $\text{delta}(s) = F(\text{reward}) - G(\text{nonreward})$, where F is some positive function of the expected probability of reward and G is some positive function of the expected probability of nonreward. Because our dependent variable is choice proportion, to the extent that rewards for responses on the left increase left responses, they must also reduce the probability of responses on the right, and conversely. If the dependent variable is right-choice proportion, s (i.e., R/(R + L)), therefore, the effects of L-rewards add in with a negative sign, L-nonrewards with a positive sign.

In the two-choice situation there are obviously four possibilities: reward and nonreward on L and R. For the two-armed bandit with payoff probabilities p (on the right) and q (on the left), the expectations are shown as:

	Outcome	
Response on left	$(1 - s)q$	$(1 - s)(1 - q)$
Response on right	sp	$s(1 - p)$
	Reward	Nonreward

Functions F and G are defined very simply:

	Outcome	
Response on left	$-\mathbf{a}(s)$	$\mathbf{b}(s)$
Response on right	$\mathbf{a}(s)$	$-\mathbf{b}(s)$
	Reward	Nonreward

The idea is simply that each reward increments s by an amount $\mathbf{a}(s)$ (i.e., by an amount that is some function of the current choice preference) and nonreward decrements s by a generally smaller amount $\mathbf{b}(s)$. $\mathbf{a}(s)$ and $\mathbf{b}(s)$ must obviously be such as to limit s to the range 0–1.

We term the assumption that the absolute magnitudes of the changes in s due to reward and nonreward are independent of the source (a left or right response) *source independence.* It is this feature that distinguishes this model from linear-operator models such as those of Bush and Mosteller (1955; see Horner & Staddon, 1987, for more detailed comparisons).

Combining the quantities in the boxes above and simplifying yields the following expression for delta(s), the expected change in s, in the two-armed bandit situation:

$$\text{delta}(s) = s[(\mathbf{a}(s) + \mathbf{b}(s))(p + q) - 2\mathbf{b}(s)] + \mathbf{b}(s) - q[\mathbf{a}(s) + \mathbf{b}(s)]. \quad (9)$$

In words, Equation 9 simply says that delta(s) is the sum of the probabilities of the four outcomes each multiplied by the appropriate change in s ($\mathbf{a}(s)$ or $\mathbf{b}(s)$), with the appropriate sign (positive for reward on R and nonreward on L, negative for reward on L and nonreward on R).

Without knowing the forms for the reward and nonreward increments, $\mathbf{a}(s)$ and $\mathbf{b}(s)$, little can be deduced from Equation 9. But if we admit the constraint that the ratio of reward and nonreward effects is constant (we term this assumption *ratio invariance*), Equation 9 can easily be reduced to a useful form. If $\mathbf{a}(s)/\mathbf{b}(s)$ is constant, then so is a quantity we term the *effect ratio,* $w = \mathbf{b}(s)/[\mathbf{a}(s) + \mathbf{b}(s)]$. Dividing Equation 9 by the quantity $\mathbf{a}(s) + \mathbf{b}(s)$ and making the appropriate substitutions yields

$$[1/(\mathbf{a}(s) + \mathbf{b}(s))]\text{delta}(s) = s(p + q - 2w) + w - q.$$

(Note that the positive multiplier, $\mathbf{a}(s) + \mathbf{b}(s)$, has the property of a window, as in the melioration discussion. If this quantity is large, then the model acts as if p, q, and s are computed over a small number of previous responses; if it is small, the model acts as if they are computed over a large number of responses.) Setting the multiplier equal to unity for reasons already given yields the following quasi-dynamic model:

$$\text{delta}(s) = s(p + q - 2w) + w - q. \quad (10)$$

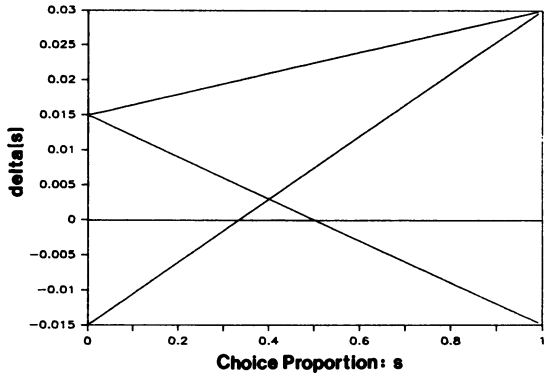

Fig. 2. Predictions of ratio invariance for the two-armed bandit problem: expected change in choice proportion, delta(s), versus choice proportion (s = R/(R + L)). Upper line: $p = 2w$, $q = w/2$: no equilibrium—prediction: exclusive choice of R; positive-slope line through abscissa: $p = 2w$, $q = 3w/2$: $s = 2/3$ is unstable equilibrium—prediction: exclusive choice of L or R, depending on starting point; negative-slope line through center: $p = q = w/2$: $s = 1/2$ is stable equilibrium—prediction: partial preference at indifference. For this graph, $w = 0.03$.

We can use Equation 10 to make predictions about the experimental situations discussed in the previous section.

Two-armed bandit. Consider first the two-armed bandit situation. Equation 10 is linear in s. Equating to zero yields the solution

$$\hat{s} = (q - w)/(p + q - 2w), \quad (11)$$

which must lie in the interval $0 \leq s \leq 1$ if the equilibrium is to be realizable, so that $p \leq w$ (and, by symmetry, $q \leq w$). If p and q are both $<w$, then the slope of the right-hand side of Equation 10 is negative, so that this partial preference is stable. If p or q are $>w$, then the only solution is exclusive choice of R or L; if both are $>w$, then exclusive choice of either L or R is possible, depending upon the initial value of s. If $p \geq q$, these three (which define the possible limit points of the theory in this situation) are the only possibilities. They are summarized in Figure 2, which shows delta(s) functions corresponding to the three possibilities.

These predictions determine the possible locations of what might be termed *preference modes,* which are measured as follows: Imagine that we record for each choice the prevailing choice proportion, computed over that choice and the $M - 1$ preceding choices. (I use the symbol S for this value, but it is important to realize that S here refers to an *empirical* quantity. In discussions of models, s

refers to a theoretical quantity: the animal's current preference. Whether or not S is in general the same as, or at least an unbiased estimate of, s depends on the animal's time window, which is not usually known. Nevertheless, there seems to be no harm in treating the two as in general equivalent, as long as the difference between them is recognized.) For example, if M is 16, then if the most recent choice is R, and the preceding 15 are 11 R and 4 L, the current value for choice proportion, S, is 0.75. The distribution of S-values across an experimental session or, more usually, a block of sessions, is the dependent variable for our analysis.

Ratio invariance then makes three kinds of predictions for the two-armed bandit situation: either a partial preference (if $p, q < w$), a single exclusive-choice mode on the majority side (if $p > w > q$), or possible exclusive-choice modes on either or both sides (if $p, q > w$).

It is important to be aware here of the limits of a quasi-dynamic analysis, which can tell where (i.e., at what s-value) a preference mode might occur, but cannot tell us whether it will occur there, or at some other permissible location, or at both. Nor can it tell us the variance of a preference distribution, or how long the process will take to stabilize (or how big a sample of data we need to be sure that all possible modes have been populated—in practice, this does not seem to be a serious problem). Despite these limitations, such models are easily disproved by persistence of a mode at an impermissible location. Absence of a mode at a permitted location is uninformative, however.

Asymmetrical frequency-dependent ratio schedule. As a second example, consider the AFD schedule discussed earlier. Recall that in this procedure, p and q are functions of (measured) choice proportion, S. In this case, $p(S) = kS$, and $q(S) = 2kS$ (i.e., payoff probability increases for both responses as the proportion of R responses increases) but payoff probability on the left is always twice that on the right. Substituting kS for p and $2kS$ for q in Equation 10 and simplifying yields

$$\text{delta}(s) = 3kS^2 - 2S(w + k) + w. \quad (12)$$

The two roots of this equation represent potential equilibria. Horner and Staddon (1987) show that it permits only two possibilities: (a) a partial-preference mode that must always be in the region $0 < S < ½$, and (b) a possible second mode, at exclusive minority choice ($S = 1$), when the partial-preference mode is at an S-value less than ⅓.

Concurrent VI VI schedule. To derive predictions, we need to arrive at an expression for p and q as a function of the relative amount of responding, s and $1 - s$, allocated to right and left. (I used t, relative time allocation, as the dependent variable in the discussion of melioration, but with the simplifications I have made, t (relative time allocation) and s (relative response allocation) have exactly the same properties on concurrent VI VI schedules. I use s here for continuity with the other published discussions of ratio invariance.) With the same assumptions as before (constant overall reward rate, constant overall response rate, response rate and allocation sufficient to make obtained reward rates equal to scheduled) we can simplify the molar feedback function for random responding (Staddon & Motheral, 1978) to arrive at a reasonable approximation: $R(x) = ax/(a + x)$, where $R(x)$ is obtained reward rate, x is (overall, not local) response rate on the right, and a is VI rate (1/VI value), as before. If $x \gg a$ (i.e., high response rate) we can neglect a in the denominator, so that $R(x) = a$. Because reward probability, p, is just $R(x)/x$, we arrive at $p = a/x$ as a reasonable approximation. If overall response rate is constant and equal to C, then $p = (a/C)/(x/C)$; but $x/C = s$, the proportion of right responses, so if we define a new variable, A, the ratio of scheduled reward rate to total response rate, $A = a/C$, then $p = A/s$ and, by symmetry, $q = B/(1 - s)$. Substituting in Equation 10 yields

$$\begin{aligned}\text{delta}(s) &= s(A/s + B/(1 - s) - 2w) \\ &\quad + w - B/(1 - s), \\ &= A - B - 2ws + w, \end{aligned} \quad (13)$$

which has a stable root at

$$\hat{s} = (A - B + w)/2w, \quad (14)$$

which is within the closed interval 0–1 if $-w \leq A - B \leq w$.

Equation 14 approximates matching when the difference between scheduled reward rates is low and response rates are relatively high (which is usually the case on VI schedules). For example, suppose that $A + B$ is constant, K (which will be true if VIs are chosen so that total reward rate is approximately constant and overall response rate is also constant). If

K is approximately equal to w, then Equation 14 becomes $s = (A - B + A + B)/2(A + B) = A/(A + B) = a/(a + b)$ (i.e., perfect matching). From Equation 14 it is easy to see that as K (total reward rate) decreases below w (so that $A - B$ also decreases if A and B are in constant ratio), the relation between s and $a/(a + b)$ approaches indifference (i.e., increasing undermatching): as $K \to 0$, $s \to \frac{1}{2}$. Conversely, if $A + B > w$, Equation 14 predicts overmatching. Perfect matching is therefore to be expected only at an intermediate range of reward frequencies.

Concurrent VI VR schedule. Again, I assume VI on the right, random-ratio (VR) on the left. Equation 9 cannot be directly applied here, because it assumes symmetry between the two choices (i.e., source independence: the idea that reward or nonreward have the same absolute-magnitude effect on choice proportion, independently of which response produced them). Because animals typically respond faster, with a different topography, on the ratio than the interval alternative, adapting to temporal differences between the schedules (reward probability increases with interresponse time on VI, but not on VR), it is unlikely that R and N will have the same effects on both alternatives in the concurrent VI VR situation. Reward for a ratio response is likely to have a different effect (call it $\mathbf{A}(s)$) on choice proportion than will reward for an interval response ($\mathbf{a}(s)$), and similarly for the effects of nonreward. With these changes, the delta(s) function (analogous to Equation 9) becomes

$$\begin{aligned}\text{delta}(s) = s[&p(\mathbf{a}(s) + \mathbf{b}(s)) + q(\mathbf{A}(s) + \mathbf{B}(s)) \\ &- (\mathbf{b}(s) + \mathbf{B}(s))] \\ &- q(\mathbf{A}(s) + \mathbf{B}(s)) + \mathbf{B}(s). \qquad (15)\end{aligned}$$

Equation 15 cannot be simplified as easily as Equation 9. Yet even if $\mathbf{a}(s) \neq \mathbf{A}(s)$, it is not unreasonable to assume that $\mathbf{a}(s) = h\mathbf{A}(s)$ (and $\mathbf{b}(s) = h\mathbf{B}(s)$), where h is a positive constant less than one that represents the slope of the linear constraint that typically relates the response rates on the two alternatives (see below). If we again assume ratio invariance, so that $\mathbf{b}(s)/[\mathbf{a}(s) + \mathbf{b}(s)] = \mathbf{B}(s)/[\mathbf{A}(s) + \mathbf{B}(s)] = w$, then, after dispensing with positive multipliers as before, Equation 15 reduces to

$$\text{delta}(s) = s[p + hq - 2w)] + h(w - q). \quad (16)$$

From the earlier discussion of concurrent VI VI, we can assume that $p = a/x$, where x is overall (not local) response rate. From published data on concurrent schedules we can also assume that rates of responding to the two alternatives are constrained by the linear relation $x + hy = K$, where $0 < h < 1$ (i.e., response rate to the ratio side, y, is generally higher, cf. Bacotti, 1977); given that $s = x/(x + y)$, we can eliminate s from Equation 16 to yield the following stable solution for x as a function of q and a:

$$\hat{x} = [a - K(q - w)]/[w(3 - h) - q] \quad (17)$$

Equation 17 is linear in a if q, reward probability on the left, is held constant. We can easily see if it is compatible with matching by varying a, obtaining x from Equation 17 and y from the linear constraint, and then plotting s against relative reward rate, $a/(a + qy)$. A typical function is shown in Figure 3: it is strongly nonlinear, and biased in favor of the ratio alternative for most values of reward proportion. A plot of the logarithms of reinforcement ratios versus response ratios (i.e., $\ln(\text{R}(x)/\text{R}(y))$ vs. $\ln(x/y)$) is also strongly nonlinear.

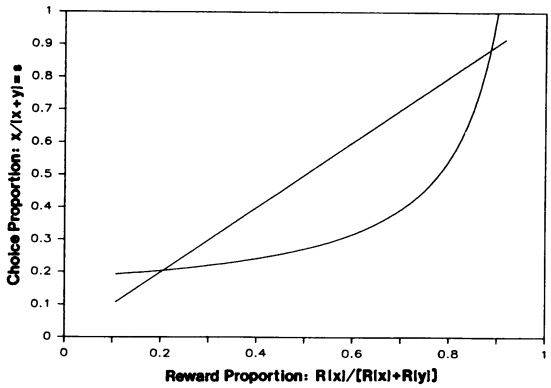

Fig. 3. Matching predictions of ratio invariance for concurrent VI VR schedules. Straight line is simple matching (proportion of VI responses equal to proportion of VI rewards). Curved line is the relation predicted by the asymmetrical version of ratio invariance in the text. Ratio value was held constant at 100 (q = 0.01), and VI rate (a = 1/VI value) was varied over the range 0.1–5.05. The other parameter values (see text) were: h = 0.5, w = 0.03, K = 100.

Frequency-dependent VI schedule. The predictions for this procedure can be derived from the VI analysis, above. Because the VI schedule values in this procedure change depending on the prevailing preference (s-value), the delta(s) function, based on Equation 13, will have three different regions. The feedback function for this schedule is defined by two values of the VI parameter:

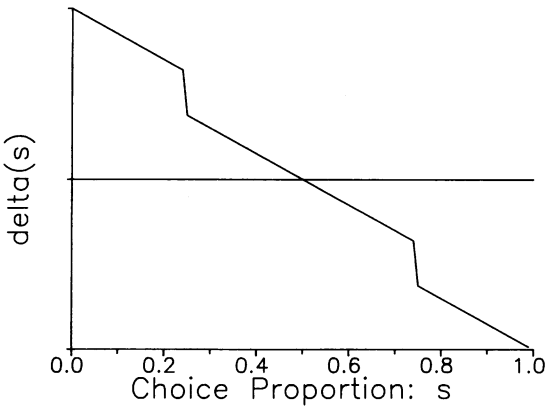

Fig. 4. Predictions of ratio invariance for the interdependent concurrent VI VI schedules shown in the top panel of Figure 1: delta(s) versus choice proportion (s = R/(R + L)). Changing the relative values of the two VI schedules affects only the height of the left and right "limbs" of the delta(s) function, which have no bearing on the equilibrium point. Indifference (s = 0.5) is always the only stable equilibrium.

$$r_R(s) = a/s, \quad \text{for } s < s_{Rcrit}$$
$$r_R(s) = b/s, \quad \text{for } s > s_{Rcrit},$$

where a and b are maximum-scheduled reward rates (i.e., 1/VI values), $a > b$, and s_{Rcrit} = ¼, and similarly for $r_L(s)$. That is, when responding is predominantly on the right ($s > s_{Rcrit}$), there is a low VI schedule for R responses and a high VI for L responses; when responding is predominantly on the left ($s < 1 - s_{Rcrit}$) there is a low VI schedule for L responses and a high VI for R. Otherwise (in the middle region, when $s_{Rcrit} < s < 1 - s_{Rcrit}$), the same low VI operates for both choices.

Figure 4 shows a plot of the three-part ratio-invariance delta(s) function for the Silberberg-Ziriax frequency-dependent schedule. It is similar to the melioration function for this procedure, shown in the two bottom panels of Figure 1: there are discontinuities at the two region boundaries, and a stable partial preference at indifference.

Comparisons with Data

Two-armed bandit. Horner and Staddon (1987) report and review data showing (a) that under some conditions, with equal payoff probabilities ($p = q$), higher absolute values for p lead to exclusive choice of one alternative, low values to indifference (i.e., a partial preference); (b) animals will often fixate, if not permanently at least for long periods, on the minority (lower probability alternative); that is, there seem to be preference modes at both exclusive-choice options under many conditions. (c) most commonly, with $p \neq q$, animals show a single preference mode at exclusive choice of the majority (i.e., higher probability) alternative. Although the necessary and sufficient conditions for these various patterns are yet to be defined, the existence of three possible patterns—exclusive choice of the majority and, occasionally, minority alternative, and partial preference—seems to be reasonably well established.

Melioration predicts reliable choice of the majority alternative and no effect of absolute reward probability. Melioration never predicts a partial preference (as long as the two probabilities are unequal), or any systematic choice of the minority alternative. Ratio invariance, however, predicts all the patterns that have been observed. For example, when the effect ratio, w, the relative effect of nonreward and reward, is greater than both p and q, ratio invariance predicts a partial preference—indifference if $p = q$. If $p, q > w$, the animal may show a preference mode at the minority alternative as well as at the majority. Ratio invariance predicts reliable, exclusive majority choice only when $q < w < p$.

Concurrent VI VI. On concurrent variable-interval schedules, animals very often vary the spacing between responses to take account of the way that payoff probability changes with postresponse time (momentary maximizing: Hinson & Staddon, 1983; Shimp, 1966). It is not clear how momentary maximizing interacts with other processes, such as melioration or ratio invariance. Hence, a failure of either theory to agree with data from choice procedures involving VI schedules may reflect inadequacy of the theory or some unknown effect of momentary maximizing. Nevertheless, it is of interest to see how well our two theories do in predicting choice between VI schedules, even if lack of correspondence between theory and data is at present hard to interpret.

Melioration predicts simple matching of response ratios to obtained reward ratios: $x/y = R(x)/R(y)$, which is a common outcome in these experiments. Not infrequently, however, there are deviations in the direction of undermatching: $x/y = [R(x)/R(y)]^v$, $0 < v < 1$ (cf. review in Baum, 1979) that are not predicted by melioration. Fantino, Squires, Delbrück and

Peterson (1972) have also shown a dependence of matching on the *absolute* value of the VI schedules. When the VI values are small (i.e., high scheduled reward rates), animals tend to overmatch (i.e., $v > 1$, in the power form of the matching equation) with respect to *scheduled* reward rates. Choice proportions are more extreme than scheduled reward proportions at high absolute reward rates (see also Alsop & Elliffe, 1988).

As I showed earlier, ratio invariance predicts matching as long as the sum of the reward rates, divided by response rate (scaled reward rate), is on the same order as the effect ratio w. When scaled total reward rate declines below w, however, there is increasing undermatching; when it is greater than w, there should be overmatching. Undermatching is the most common systematic deviation from matching (Baum, 1979). There are also data from Fantino et al. (1972) showing that the tendency to overmatch to *scheduled* reward ratios increases with absolute reward rate. Although the conditions under which these changes take place are not yet perfectly defined, it is apparent that ratio invariance predicts both matching and the usual deviation from matching, whereas melioration predicts matching alone. It might appear that the degree of undermatching should increase as scheduled VI rate decreases, but there is some evidence for a countervailing decrease in w as absolute reward rate decreases (cf. Horner & Staddon, 1987), so that a decrease in reward rate does not necessarily increase the difference between $|A - B|$ and w. Hence, it is not clear exactly when we should expect undermatching and overmatching, although it is clear that ratio invariance permits matching, as well as both types of systematic deviation therefrom, on concurrent VI VI schedules.

Concurrent VI VR. Because this procedure also involves VI, the same caveat about momentary maximizing applies. This complication apart, melioration predicts simple matching with respect to time but biased matching with respect to responses. Herrnstein and Heyman (1979), in a careful analysis, report biased matching with respect to both time and responses; specifically

$$s/(1 - s) = 0.718[\mathrm{R}(x)/\mathrm{R}(y)]^{1.041}$$

and

$$t/(1 - t) = 1.291[\mathrm{R}(x)/\mathrm{R}(y)]^{1.036},$$

where $\mathrm{R}(x)$ and $\mathrm{R}(y)$ are obtained reward rates on the interval and ratio sides, respectively, and t and s are time and response proportions. The exponent here is not significantly different from unity, so there is no tendency to over- or undermatch, which is consistent with melioration, as is the bias in response allocation. But simple time-allocation melioration cannot account for the time-matching bias.

Ratio invariance cannot account for any kind of matching on VI VR schedules. It predicts a strong, nonlinear bias in favor of the ratio alternative over much of the range, which is not reflected in published data.

It is worth noting that the asymmetrical version of RI that fails to account for concurrent VI VR schedules (perhaps because of the asymmetrical involvement of temporal factors) may nevertheless prove useful in application to the single-response case. For example, consider a single-response probabilistic schedule in which the two responses are pecking and not-pecking for some brief period of time; if this brief period is chosen to be short enough that no more than one response can occur, then choice proportion, s, is proportional to response rate. If the changes produced by R or N ($\mathbf{A}(s)$, $\mathbf{B}(s)$) are larger for pecking than for not-pecking (i.e., $\mathbf{A}(s) \gg \mathbf{a}(s)$, etc.), then when pecking is no longer reinforced, it will decline to a negligible level. The argument is as follows: From Equation 16, the equilibrium value of choice proportion, $\hat{s}$, is

$$\hat{s} = [h(q - w)]/[p + hq - 2w), \quad (18)$$

where $h = \mathbf{a}(s)/\mathbf{A}(s)$, and so on, as before. If $p = q = 0$ (i.e., extinction), then $\hat{s} = h/2$, which will be small if h is small (i.e., if $\mathbf{A}(s) \gg \mathbf{a}(s)$). Spaced-responding schedules can be looked at as a variety of frequency-dependent schedule, because the probability of reward for pecking is inversely related to its frequency and directly related to the frequency of not-pecking. It is easy to show that RI predicts quite unstable performance under these conditions, although the details of the argument would take us beyond the topic of this paper.

Asymmetric frequency-dependent ratio schedule. The results of this procedure are described in Horner and Staddon (1987): When payoff probability on the left is twice that on the right, and both increase linearly with the proportion of right choices (i.e., with S), the usual result is a preference mode on the majority (left) side.

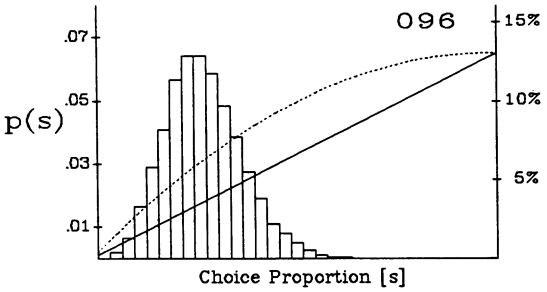

Fig. 5. Choice proportion (S = R/(R + L)), averaged in a moving window of 32 responses across 10 experimental sessions for a single pigeon, on the asymmetrical interdependent schedule described in the text. The solid line shows the payoff probability on the right key (payoff on the left was always twice this). The dotted line shows the overall reward probability, which is maximal at exclusive choice of the right key (from Horner & Staddon, 1987, Figure 7).

Some animals, with modes far to the left, occasionally show a second mode at exclusive minority (right) choice. Both results are consistent with ratio invariance, neither with melioration, which predicts exclusive majority (left) choice (hence, a very low payoff probability). Figure 5 shows typical data from one pigeon.

Frequency-dependent VI schedule (Silberberg & Ziriax, 1985). Both melioration and ratio invariance predict a stable equilibrium at indifference (s = 0.5) on this procedure, which rewards both responses according to the same VI schedule in the middle range of s-values, but rewards the less frequent response at a higher level at extreme s-values. Silberberg and Ziriax report this result—indifference—when the t-value (time proportion) is computed with a long time window (4 min), but report modes at both extremes when they used a short (6-s) time window (their Figure 4). Their simulation of molecular maximizing shows similar results: indifference when the time window is long, modes at the extremes when it is short.

The Silberberg-Ziriax results raise an interesting methodological point, which is noted by Vaughan (1987) and illustrated with data, but can also be made a priori: The preference distribution must depend to some extent on the time or response window on which it is based. For example, in the limit, if we use a window so short that no more than one response can occur within, the choice distribution will necessarily be bimodal, because only S-values of 0 or 1 are possible. In other words, if the time window is shorter than the typical "stay" time (interchangeover time), then a bimodal choice distribution will likely result, no matter what the animal's "true" preference. It is possible to argue, therefore, that Silberberg and Ziriax's finding, with a short time window, of a bimodal choice distribution is an artifact, and that their data really show a mode at indifference at both time windows, a result consistent with both melioration and ratio invariance. This argument is also consistent with Silberberg and Ziriax's finding of a negative correlation between successive time proportions at the short time window (their Figures 5 and 7). Suppose, for example, that animals switch at a roughly constant rate, with a typical "stay time" slightly longer than the averaging window. Then a high time proportion (t-value) will mean that the averaging window almost coincided with a stay on the right, so that the following window must almost coincide with a stay on the left (low t-value). Hence, high t-values will invariably be followed by low ones, intermediate values will consistently be followed by intermediate ones, and so on, yielding a linear function with negative slope slightly less than unity, similar to the ones reported. Using a moving window rather than a block-by-block window eliminates this aliasing problem and much reduces the bimodality artifact, which is further reduced by basing the window on response count, rather than time, so as to maintain the same sample size for every measurement. The data shown in Figure 5 were obtained in this way: Note that if this artifact was operative there, we would expect to see a mode at least at an S-value of zero (i.e., on the left) where most S-values lie. Hence the absence of a mode on the left gives us confidence that these data are not artifactual.

This problem is not solved by assuming that the entire session duration is the appropriate time window—this decision is as arbitrary as any other. The only real solution is a true dynamic model, whose properties would tell us the proper window. Lacking that, we are forced back to generalities, namely to judge any theory by its ability to explain a body of data gathered in some consistent and reproducible way.

Conclusion

In this review I have distinguished two main types of theory, static and dynamic, and identified a third type, which I term quasi-dy-

namic. Quasi-dynamic theories are dynamic in form, that is, they define the expected change in behavior at a given point, but they are specified only to the level that they permit predictions of stable and unstable equilibria, not to the point where they permit true dynamic predictions of learning curves and the like.

It turns out to be fairly straightforward to describe the hypothesis of melioration in a quasi-dynamic form, and to use this formulation to derive predictions for five kinds of choice situations: concurrent VI VI schedules, concurrent VI VR schedules, the two-armed bandit (concurrent random-ratio schedules), frequency-dependent VI VI schedules, and asymmetrical frequency-dependent random-ratio schedules. Melioration predicts reasonably well only the matching data from the two interval procedures. Yet even there it fails to predict most of the orderly departures from matching, such as undermatching and overmatching to scheduled values on concurrent VI VI schedules. Melioration predicts the matching outcome on concurrent VI VR and the typical response bias towards the ratio side, but fails to predict the time bias towards the interval side.

Melioration predicts only one of the three patterns that have been reported in the two-armed bandit situation, and fails completely to predict performance on the asymmetrical frequency-dependent ratio schedule studied by Horner and Staddon (1987; they show that it also fails to predict performance on a symmetrical frequency-dependent ratio schedule).

Ratio invariance, the alternative quasi-dynamic model proposed by Horner and Staddon, predicts the different patterns observed in the two-armed bandit situation, it predicts matching, undermatching and overmatching on concurrent variable-interval schedules, and it predicts in detail, and in a way that accommodates individual differences, the preference patterns on asymmetrical (and symmetrical) frequency-dependent ratio schedules. Ratio invariance clearly fails only in the VI VR case, which violates its assumption of symmetry, and involves temporal factors that are explicitly excluded from the theory.

The major uncertainty about ratio invariance is the factors that determine w, the effect ratio, because the behavior of the model often depends critically on it. For example, it is essential that w be appropriately related to typical reward-probability values for correct predictions of concurrent VI VI performance. Yet there is nothing in the theory as it stands that requires w to be in the appropriate range. Nevertheless, in situations specifically designed to test the theory in ways that do not depend upon particular values for w, and in direct competition with other accounts such as melioration and molar and momentary maximizing, it has so far been strongly confirmed. It seems fair to conclude that ratio invariance is a more accurate model than its competitors to explain how pigeons adapt in the short term to purely probabilistic situations, but that other processes—momentary maximizing, melioration—come in to play when temporal factors are involved, especially when the situation is asymmetrical. I do not make much of the ability of RI to make good predictions in the symmetrical concurrent VI VI case, because almost any reward-following process is adequate to explain matching under these conditions (cf. Hinson & Staddon, 1983; Staddon et al., 1981).

The melioration hypothesis was originally derived by assuming that matching at the molar (whole-session) level is a consequence of a matching-like process at the molecular (second-by-second) level. This kind of reasoning can be dangerous, because regularity at the molar level may reflect any one (or more) of a number of local processes. The larger the aggregate that enters into any lawful relationship, the less sure we can be of the local process or processes that underlie it—there is a convergence from the molecular to the molar. This is demonstrably true for concurrent-interval schedules, in which several theoretical studies have shown that a variety of choice rules, from momentary maximizing to reward following, produce molar matching. The present review shows that reasoning directly from the molar to the molecular may have failed in the case of melioration, because the hypothesis is probably false as a universal model for free-operant choice, at least in the simple form tested here.

The validity of this conclusion depends, of course, on the validity of my quasi-dynamic version of melioration, which is admittedly highly simplified, especially in terms of the local feedback functions I have used for variable-interval schedules. Simplified or not, the model is adequate to arrive at the correct prediction for concurrent VI VI schedules, the situation for which, and from which, melioration was originally derived. Equally clearly, it fails in situations such as the asymmetrical and

symmetrical frequency-dependent ratio schedules, where there can be little doubt about the local feedback function.

Ratio invariance has a number of attractive features as an elementary reward-following process. It is truer to the original law of effect than comparative principles such as matching and melioration that are nominally derived from it (Herrnstein, 1970) because unlike melioration (for example) it includes a real "strengthening" principle: individual rewards actually increase the probability of the rewarded action, and nonrewards decrease its probability. A purely comparative rule like melioration is silent on the effect of particular rewards and nonrewards and deals only with some property of their aggregate difference.

The very indefiniteness of RI is an asset because it predicts only those features of the data, steady-state patterns, that seem to be orderly, and is noncommittal about transitional patterns and details of sequential relationships that seem to be highly variable.

Ratio invariance has some resemblance to the old idea of a *reflex reserve* (Skinner, 1938), with the effect ratio, w, having some of the properties of the *extinction ratio*. The extinction ratio is the number of responses (in a single-response situation) that are generated by a single reinforced instance. In practice, this number does not seem to be constant so the idea was abandoned, but the idea that responding on ratio schedules typically changes its form above a certain ratio value (ratio "strain") does have some factual basis. The parallel with w can be seen if we extend the tentative development I gave earlier for the single-response case (i.e., the two-response situation with the second response a brief period of nonresponding that is never reinforced). Equation 16, with $q = 0$, becomes

$$\text{delta}(s) = s[p - 2w)] + hw,$$

which implies exclusive choice (slope positive) when $p > 2w$, and partial preference, with $\hat{s} = -hw/(p - 2w)$, otherwise. Thus, the ratio of size $1/2w$ has properties similar to the old extinction ratio although the transition implied is not from responding to extinction but from exclusive choice (normal, high-rate responding) to partial preference ("strain").

Ratio invariance is more adaptive than hill-climbing processes such as melioration and momentary maximizing because it is less susceptible to obvious "traps." A simple hill-climber that finds itself in a poor situation that nevertheless contains relatively good and bad alternatives is constrained to fixate on the best of this particular bad lot. This is because hill-climbing only compares alternatives with each other, not with an internal standard. Ratio invariance, on the other hand, compares alternatives both with each other and with an internal standard, and does so in the simplest possible way. An RI-type organism in a poor situation (i.e., no payoff probability $>w$) will devote more responding to the better alternatives, but will continue to sample all, and so will detect a change for the better should one occur. Conversely, in a rich situation, an RI organism may fixate on any alternative. RI is a process that satisfies (Simon, 1956) rather than optimizes a sensible strategy in an uncertain world where achieving the best is often less important than avoiding the worst.

REFERENCES

Alsop, B., & Elliffe, D. (1988). Concurrent-schedule performance: Effects of relative and overall reinforcer rate. *Journal of the Experimental Analysis of Behavior,* **49,** 21–36.

Bacotti, A. V. (1977). Matching under concurrent fixed-ratio variable-interval schedules of food presentation. *Journal of the Experimental Analysis of Behavior,* **27,** 171–182.

Baum, W. M. (1979). Matching, undermatching, and overmatching in studies of choice. *Journal of the Experimental Analysis of Behavior,* **32,** 269–281.

Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning.* New York: Wiley.

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

Crutchfield, J. P., Farmer, J. D., Packard, N. H., & Shaw, R. S. (1986). Chaos. *Scientific American,* **255**(6), 46–57.

Fantino, E., Squires, N., Delbrück, N., & Peterson, C. (1972). Choice behavior and the accessibility of the reinforcer. *Journal of the Experimental Analysis of Behavior,* **18,** 35–43.

Fisher, R. A. (1930). *The genetical theory of natural selection.* Oxford: Clarendon Press.

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

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

Herrnstein, R. J., & Heyman, G. M. (1979). Is match-

ing compatible with reinforcement maximization on concurrent variable interval, variable ratio? *Journal of the Experimental Analysis of Behavior,* **31,** 209–223.

Herrnstein, R. J., & Vaughan, W., Jr. (1980). Melioration and behavioral allocation. In J. E. R. Staddon (Ed.), *Limits to action: The allocation of individual behavior* (pp. 143–176). New York: Academic Press.

Heyman, G. M. (1979). Matching and maximizing in concurrent schedules. *Psychological Review, 86,* 496–500.

Hinson, J. M., & Staddon, J. E. R. (1983). Matching, maximizing, and hill-climbing. *Journal of the Experimental Analysis of Behavior,* **40,** 321–331.

Horner, J. M., & Staddon, J. E. R. (1985, November). *Choice on probabilistic schedules: A reward-following analysis.* Paper presented at the annual meeting of the Psychonomic Society, Boston.

Horner, J. M., & Staddon, J. E. R. (1987). Probabilistic choice: A simple invariance. *Behavioural Processes,* **15,** 59–92.

Lea, S. E. G. (1981). Concurrent fixed-ratio schedules for different reinforcers: A general theory. In C. M. Bradshaw, E. Szabadi, & C. F. Lowe (Eds.), *Quantification of steady-state operant behaviour* (pp. 101–112). Amsterdam: Elsevier/North-Holland Biomedical Press.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature,* **261,** 459–467.

Maynard Smith, J. (1982). *Evolution and the theory of games.* Cambridge: Cambridge University Press.

Myerson, J., & Hale, S. (1984). *Transition-state behavior on conc VR VR: A comparison of melioration and the kinetic model.* Paper presented at the Meeting of the Association for Behavior Analysis, Nashville, TN.

Myerson, J., & Miezin, F. M. (1980). The kinetics of choice: An operant systems analysis. *Psychological Review,* **87,** 160–174.

Rachlin, H. (1973). Contrast and matching. *Psychological Review,* **80,** 217–234.

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

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

Richerson, P. J., & Boyd, R. (1987). Simple models of complex phenomena: The case of cultural evolution. In J. Dupré (Ed.), *The latest on the best: Essays on evolution and optimality* (pp. 27–52). Cambridge, MA: Bradford/MIT Press.

Schelling, T. C. (1978). *Micromotives and macrobehavior.* New York: Norton.

Shimp, C. P. (1966). Probabilistically reinforced choice behavior in pigeons. *Journal of the Experimental Analysis of Behavior,* **9,** 443–455.

Silberberg, A., & Ziriax, J. M. (1985). Molecular maximizing characterizes choice on Vaughan's (1981) procedure. *Journal of the Experimental Analysis of Behavior,* **43,** 83–96.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review,* **63,** 129–138.

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

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

Staddon, J. E. R. (1982). Behavioral competition, contrast and matching. In M. L. Commons, R. J. Herrnstein, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 2. Matching and maximizing accounts* (pp. 243–261). Cambridge, MA: Ballinger.

Staddon, J. E. R., Hinson, J. M., & Kram, R. (1981). Optimal choice. *Journal of the Experimental Analysis of Behavior,* **35,** 397–412.

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

Staddon, J. E. R., & Motheral, S. (1979). Response independence, matching, and maximizing: A reply to Heyman. *Psychological Review,* **86,** 501–505.

Sternberg, S. (1963). Stochastic learning theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 1–120). New York: Wiley.

Vaughan, W., Jr. (1981). Melioration, matching, and maximization. *Journal of the Experimental Analysis of Behavior,* **36,** 141–149.

Vaughan, W., Jr. (1982). Choice and the Rescorla-Wagner model. In M. L. Commons, R. J. Herrnstein, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 2. Matching and maximizing accounts* (pp. 263–279). Cambridge, MA: Ballinger.

Vaughan, W., Jr. (1985). Choice: A local analysis. *Journal of the Experimental Analysis of Behavior,* **43,** 383–405.

Vaughan, W., Jr. (1987). Reply to Silberberg and Ziriax. *Journal of the Experimental Analysis of Behavior,* **48,** 333–340.

Weisbuch, G. (1986). Networks of automata and biological organization. *Journal of Theoretical Biology,* **121,** 255–267.

Received April 13, 1987
Final acceptance November 1, 1987

APPENDIX

Definitions for Symbols Used in the Text

s = probability of a response on the right (a theoretical variable).

S = proportion of right responses, measured in a moving window of size M (empirical variable).

M = window size.

delta(s) = expected value for the change in s associated with a given s-value and known schedule parameters.

a, b = scheduled VI reinforcement rates (=1/VI value) on the right and left.

A, B = scheduled VI reinforcement rates on the right and left, scaled in terms of overall response rate, $C = x + y$: $A = a/C$, $B = b/C$.

p, q; $p(S)$, $q(S)$ = reinforcement probabilities on the right and left.

x, y = overall response rates on the right and left.

$\mathrm{R}(x)$, $\mathrm{R}(y)$ = obtained overall reinforcement rates on the right and left.

x_{R}, y_{L} = local response rates on the right and left.

t = proportion of time spent responding on the right (theoretical variable).

r_{R}, r_{L}; $r_{\mathrm{R}}(t)$, $r_{\mathrm{L}}(1 - t)$; $r_{\mathrm{R}}(x_{\mathrm{R}})$, $r_{\mathrm{L}}(y_{\mathrm{L}})$ = local reinforcement rates on the right and left.

delta(t) = expected value for the change in t associated with a given t-value and known schedule parameters.

$\mathbf{a}(s)$, $\mathbf{b}(s)$; $\mathbf{A}(s)$, $\mathbf{B}(s)$; $\mathbf{a}(t)$, $\mathbf{b}(t)$ = increments or decrements in response proportion or proportion of time spent caused by reward or nonreward according to a reward-following model.