7 How Each Reinforcer Contributes to Value: "Noise" Must Reduce Reinforcer Value Hyperbolically

Michael L. Commons
Harvard Medical School

Michael Woodford
University of Chicago Business School

Edward J. Trudeau
Harvard University

Empirical evidence has shown that the effectiveness of a reinforcer is related to the delay between reinforcer and response (e.g., for pigeons, see Chung & Herrnstein, 1967; Mazur, 1987). For people the finding of a similar function shows that the effectiveness of a stimulus is related to the delay between its presentation and the time at which it is to be recalled (for people, see Ebbinghaus, 1885). This decrease in effectiveness of an event over time has come to be known as the recency effect (e.g., Glanzer & Cunitz, 1966; Madigan, 1971; Murdock, 1962; Sternberg, 1966; Woodword, 1970). That reinforcer effectiveness over time may be represented as an exponential decay was not an unreasonable hypothesis, and has been suggested previously (e.g., Commons, 1981, and White & McKenzie, 1982, in animal psychology; Muth, 1960, in forecasting theory; Wickelgren, 1974, in human psychology). Yet previous data suggest that this relation is hyperbolic (Commons, Woodford, & Ducheny, 1982; McCarthy's data as analyzed in McCarthy & White, 1987). Theoretical considerations from a number of sources suggest that the decay of value is best represented by a hyperbolic model (Ainslie, 1975; Davison & Tustin, 1978; Fantino, 1981; Fantino, Abarca, & Dunn, 1987; Mazur, 1987).

The reduction in reinforcer effectiveness implicit in the hyperbolic model may be due to the imperfections of memory and the confusion caused by intervening events. From a signal-detection perspective, intervening events function as noise under this supposition, obscuring previous events. If reinforcer effectiveness decreases hyperbolically, then the decrease will be rapid at first and fall more slowly as time progresses, though perhaps never reaching zero. It will be shown that the relationship between reinforcer effectiveness and reinforcer delay is predicted by a *hyperbolic model* derived using *statistical decision theory* (Raiffa, 1968; Raiffa & Schlaifer, 1961) and the additive noise assumption in Woodford's *additive noise model* (Commons, Woodford, & Ducheny, 1982), which is based on signal-detection theory (Egan, 1975).

THE EFFECTS OF DELAY BETWEEN ANTECEDENT REINFORCERS AND SUBSEQUENT RESPONSES

In order to assess the effect of delay between antecedent reinforcers and responses that follow them, discrimination procedures have been devised and utilized (e.g., Commons, 1979, 1981; Commons, Woodford, & Ducheny, 1982; Mandell, 1981, 1984). The reinforcement schedules used as stimuli in discrimination studies have been samples from a variety of schedules (Commons & Nevin, 1981). Mandell used rich and lean VI schedules in which the pattern and spacing of reinforcers on each trial was not measured; Commons (1979, 1981) used samples from a rich or a lean T-schedule that resembled random-interval (RI) schedules. Here, the derivations here will be carried out for a general form of Commons's (1981) procedure.

A number of questions about the nature of decrementation and the basic processes that cause decrementation have arisen. These questions will be addressed by examining decrementation functions. A theoretical basis that accounts for the hyperbolic decreases in perceived reinforcement value and in reinforcement effectiveness has not been set forth. Here, the form will be demonstrated theoretically, but not empirically, to be hyperbolic.

The purpose of this chapter is to show how the additive-noise account, which is consistent with delay data, leads to a hyperbolic decrementation function. Woodford constructed this account (see Commons, Woodford, & Ducheny, 1982) to explain the results obtained from preference- and discrimination-trial experiments. The *additive noise model* might fairly be called one of the simplest of the many models proposed thus far in that it makes only a few assumptions.

The first assumption is that each and every reinforcer obtained affects choice. The effect may be small to the point of vanishing if the time between the reinforcer and the response is very large. Second, the effects of each reinforcer are added to the effects of the other reinforcers (Commons, Woodford, & Ducheny, 1982; Mazur, 1987). Third, noise that occurs randomly during the delay between a response and a reinforcer decreases the effectiveness of that reinforcer in controlling that response. The noise randomly distorts the subject's memory of reinforcers, each distortion being statistically independent of previous distortions. These are reasonable assumptions because they are consistent with a number of traditional theories of memory mechanisms (Baddeley & Hitch, 1974; Ebbinghaus, 1885; Underwood, 1948a).

The additive noise model applies to both preference and discrimination situations. Decrementation in the discriminative and the preference situations is most likely symmetrical. Many of the laws describing one will hold for the other (Williams, 1982). The amount of conditioning is equal to a function of the product of the value remembered about the reinforcer "R" and value remembered about the reinforced response "S^{R+}" (Dickinson, 1980; Hall & Pearce, 1982; Kaye & Hall, 1982; Williams, 1982). As far as the memory effect on conditioning is concerned, the two situations are symmetrical if the combining of the remembered magnitudes of the two events that influence conditioning is commutative: $a*b = b*a$. In the discriminative case, the remembered magnitude of the two events that influence conditioning is the value remembered about R times the value remembered about S^{R+}. In the preference case the remembered magnitudes of the two events that influence conditioning is the value remembered about S^{R+} times the value remembered about R.[1] The effect would be symmetrical because the discriminative case is simply the reverse of the preference case. Hence, showing why the decrementation function is hyperbolic for either situation shows it for the other. The effect of a reinforcer is decremented as a hyperbolic function of the time between the making of the choice and the onset of a reinforcer in preference situations, and the time between the onset of a reinforcer and the making of a choice in discrimination situations. In this chapter the additive noise model will be constructed for discrimination situations.

PROCEDURE

Analogue of Procedure Used in Discrimination-of-reinforcement-density Experiments

The experimental task is like the task of discriminating from which of two randomly chosen urns, one representing a rich reinforcement schedule and

---

[1] Presumably the conditioning or the establishment of control depends on the first event being available for pairing, that is, remembered long enough to interact with the second. If memory were perfect, the initial value of an event would be maintained at the time of a second event. With noise, the initial value would move toward the mean of all possible initial values. Let us say the first event has either a value of 1 or -1, and the second event a value of 1. If the combining were multiplicative, there would be conditioning in either case, $1 \times 1 = 1$, $-1 \times 1 = -1$. With noise, if initially the value of the first event was 1 it would be less than 1 when the second event occurred. If it was -1, then when the second event occurred, it would be more than -1. If there were total forgetting, the first event would be 0 whether it were 1 or -1 initially. Hence there would be no conditioning, $0 \times 1 = 0$. In the preference situation, the first event is the response and the second is the reinforcer. In the discrimination situation the order of the events are reversed.

one a lean reinforcement schedule, a sample is drawn. In the experiment, the sample is a period from a reinforcement schedule. In the analogue, the two urns or schedules contain different proportions of reinforcing and nonreinforcing balls. The rich urn, in which the probability of reinforcement is .6, has six reinforcing balls and four nonreinforcing balls. Conversely, the lean urn has four reinforcing balls and six nonreinforcing balls. During each stimulus period the subject is allowed *n* draws with replacement from an unknown urn, where *n* equals some number from 6 to 18. After the *n*th draw a choice period follows. During this period the subject indicates whether the sampled urn was the rich or the lean reinforcement one.

Method

The central purpose of this experiment is to provide a framework for the development of the additive noise model. Three versions of the experiment are used to characterize the role that time plays between a reinforcer and a choice. Qualitative tests on data that Commons (1981) previously collected show whether the pigeons' discrimination of reinforcement density is sensitive to time, relative time, or to the number of events between the reinforcer and the choice.

In Commons's 1981 study, four pigeons were each run in one 256-trial session per day. The experiment utilized a standard Skinner-box pigeon panel with three keys. Each trial consisted of a stimulus period followed by a choice period as shown in the bottom of Fig. 7.1.

The stimuli to be discriminated consisted of samples randomly drawn from one of two modified *T*-schedules (Schoenfeld & Cole, 1972) each composed of four cycles (Commons, 1979, 1981). On each trial, a substimulus sample from one of the two *T*-schedules was presented during the stimulus period. Each substimulus sample consisted of four equal-duration cycles. As shown at the top of Fig. 7.1, the cycles were numbered so that cycle C_4 in the stimulus period occurred at the beginning of the stimulus period and was furthest, temporally, from the choice period. Conversely, cycle C_1 occurred at the end of the stimulus period and was immediately followed by the choice period. Thus, a four-cycle trial would be numbered: $C_4\ C_3\ C_2\ C_1 \rightarrow$ choice period.

At the beginning of each stimulus period, the center key was white. The first center-key peck turned off that key's light for that cycle. On each cycle, a center-key peck was ($v_i = 1$) or was not ($v_i = 0$) reinforced immediately. A reinforcer was one 20-mg pigeon pellet from an illuminated hopper. On the rich reinforcement schedule, the first center-key peck in each cycle was reinforced with a probability of .75. On the lean schedule the probability of reinforcement for the first peck in each cycle was .25.

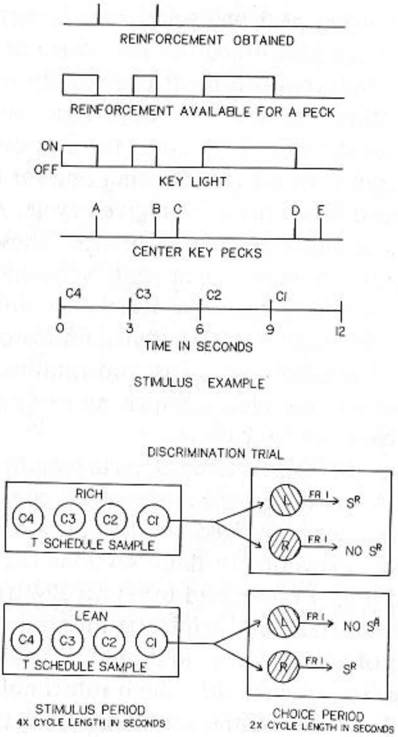

Fig. 7.1 A stimulus example from a discrimination trial shown (*top left*) by means of a state diagram for a sample from a 3-s cycle schedule to illustrate what may happen if center-key pecks occur (and do not occur) when reinforcement has been programmed (or not). In reality this would be very unusual because there is a cycle without a peck occurring in it. Contingencies during an entire trial are shown (*bottom left*). The stimulus period contains substimuli of the form shown in the top left part. The choice period, which immediately follows the stimulus period, contains a single reinforcer for correct choice or extinction for incorrect ones. There is no intertrial interval.

Rich and lean schedules were delivered with a probability of .50. In any given cycle, the first peck was reinforced with the same probability as the first peck on the rest of the cycles on that trial, $p(S^{R+}|R_c)$ having the value of either .25 for the entire trial or .75 for the entire trial. None of the center-key pecks that occurred subsequent to the first during any given cycle was reinforced.

As shown at the bottom of Fig. 7.1, after the last stimulus cycle of a given trial the choice period began. The center key was darkened and the two side keys were transilluminated, the left key with a red light, the right one with a green light. In the choice period the subject had twice the cycle length to identify the preceding stimulus period as consisting of either a sample from the rich reinforcement schedule or the lean reinforcement schedule. The first side-key peck, whether correct or not, darkened both keys, and no further pecks were counted. The rich schedule was indicated by making a left-key response, the lean schedule by a right-key response.

When a substimulus sampled from the rich schedule was presented on the center key, the first left-key peck was reinforced with a 96-mg pellet (a hit or "left-correct"). A right was not reinforced (a miss or "right-error"). When a substimulus from the lean schedule was presented on the center key, the first right-key peck was reinforced (a correct rejection or "right-correct"). A left was not reinforced (a false alarm or "left-error").

Although reinforcement probability for the first peck in any given cycle was neither one nor zero, each first center-key peck in a cycle was either reinforced or not. Therefore, for a given trial, a reinforcement pattern can be written with a *1* representing one earned reinforcer and a *0* representing no earned reinforcers for a given cycle. An obtained reinforcement pattern for a trial might be *1110* → choice. These patterns generally represent both "obtained" reinforcement and "scheduled" reinforcement [2] since subjects almost always respond at least once during each cycle (Commons, 1979). The binary pattern of scheduled reinforcement that is delivered is called the sampled substimulus or just substimulus. Here there were $2^4 = 16$ different substimulus samples, comprising every possible combination of reinforcement over the four cycles.

Three different standard cycle lengths were used: 2 s, 3 s, and 4 s ($t = 2, 3, 4$). A standard t-value remained constant across trials until the pigeon's performance stabilized so that p(L) did not vary more than +0.1 from session to session. In daily sessions the standard cycle length was of base length (i.e., t multiplied by 1) on 224 trials, doubled (t multiplied by 2) on 16 probe trials, and tripled (t multiplied by 3) on another 16 probe trials. The probe trials were randomly distributed throughout a session.

The frequencies with which substimuli for standard and probe trials with 0, 1, 2, 3, or 4 reinforcers came from the rich schedule and from the lean schedule are shown in Fig. 7.2. Any substimulus could come from either the rich or lean schedule. All that differed was the frequency with which they did so. The modal number of reinforcers was three reinforcers per trial for the rich schedule, and one for the lean schedule.

The Additive Noise Model: Derivation of a Simple Form

A model for the process by which reinforcers lose control over responses is introduced and illustrated with the four-cycle data. The general additive noise model developed here shows that a hyperbolic function must describe the decrease in value of reinforcement the further reinforcement occurs from the behavior that it controls, given some assumptions about the effect of noise on memory for reinforcement. To illustrate how the additive noise model may be applied in one experimental situation, three versions of it will be shown to distinguish qualitatively between the effects of programming a given temporal-length base cycle for a number of sessions and changing

---

[2] Only one reinforcement schedule of all the possible schedules did not appear in any trial due to the nature of the program used.

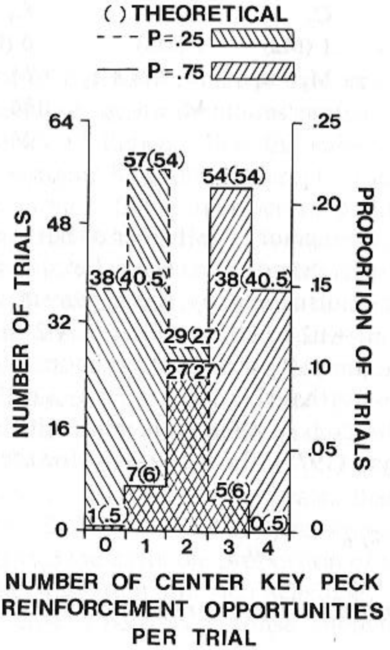

Fig. 7.2 The distribution of reinforcement for the first center-key peck ($S^{R+}|C$) for 4-cycle (*left*) are shown. The number of center-key pecks and proportion of trials on which reinforcement occurs in a sample is shown for the rich (positive-slope hatching) and lean (negative-slope hatching) schedules. The respective $p(S^{R+}|C)$ are 0.75 and 0.25. Theoretical values represent predicted outcomes based on the binomial distribution with $n = 4$ and the respective p given previously.

them momentarily on probe trials. The data will not be used to show whether the decrementation function is hyperbolic. Again, the remembered value could depend on: (a) the time between a reinforcer and choice, (b) the relative time between those events, or (c) the number of intervening cycles.

The simplest overall model of forgetting assumes that the *signal* stays constant while the amount of *noise* increases linearly with time. This is similar to the interference theories in human learning and cognition (e.g., Baddeley & Hitch, 1974; Barnes & Underwood, 1959; Ebbinghaus, 1885; Melton & Irwin, 1940; Underwood, 1948a, 1948b, 1957), as well as Grossberg's (1978) theory of scaled short term memory. The overall model assumes that at the time of reinforcement the perceived value of the reinforcement in a sample is a random variable. What is discriminated about overall reinforcement density depends on what is remembered about what happened on each cycle. Perceived value of a substimulus is a random variable because the noise added to the memory of each reinforcer makes the discrimination of that reinforcement event imperfect. The noise affects the memory of whether or not a reinforcer occurred on a particular cycle. With each cycle that occurs thereafter, another random noise component is added to the previous memory of that reinforcement. The random noise terms are independent of the initial random noise. Hence, the events on the previous cycle are remembered as follows:

C_4	C_3	C_2	C_1	→	choice
1 (M1)	1 (M2)	1 (M3)	0 (M4)		
	M_1+n_1	$M_1+n_1+n_2$	$M_1+n_1+n_2+n_3$		$M_1+n_1+n_2+n_3+n_4$
		M_2+n_2	$M_2+n_2+n_3$		$M_2+n_2+n_3+n_4$
			M_3+n_3		$M_3+n_3+n_4$
					M_4+n_4

M_1 is the memory of whether or not there was a reinforcer on cycle i and n_i represents the noise on cycle i. Let y_i be the amount of reinforcement in cycle c_i, as remembered by the organism at the time of choice. Then $y_4 =$ M1 + n1 + n2 + n3 + n4 and $y_3 =$ M2 + n2 + n3 + n4, and so on. Assume that y_i is a normally distributed random variable with variance σ^2_i and mean μ_i in the case that $vi = 1$ and mean $-\mu_i$ in the case that $v_i = 0$. Given two normal distributions of reinforcement density with mean μ_n and μ_{sn}, then according to Egan (1975), the z_{sn} – score for the sn distribution is

$$z_{s\ n} = \frac{c - \mu_{sn}}{\sigma_{sn}}\ , \tag{1}$$

and the z_n – score for the *n* distribution is

$$z_n = \frac{c - \mu_n}{\sigma_n}\ , \tag{2}$$

where c is the cutoff or criterion value along the x-axis for both distributions (not to be confused with c_i, which represents cycle i). The criterion value is that value along the decision axis above which a signal is present and below which no signal is present. The measure of overlap between the two distributions is denoted as d' and is equivalent to

$$d' = z_n - z_{sn} = \frac{\mu_{sn} - \mu_n}{\sigma_{sn}}\ . \tag{3}$$

Therefore, assuming that the two means are additive inverses involves no loss of generality, and only the difference $d' = [\mu_i - (-\mu_i)]/\sigma_i$ turns out to have observable significance. Here, the term μ_i represents the mean of the μ_{sn} distribution, or the distribution with reinforcement, and the term $-\mu_i$ is equivalent to the μ_n distribution, or the distribution without reinforcement. σ_i is the variance of the both distributions, as both are assumed to be normally distributed y_i curves. The sensitivity, *d'*, is the standard psychophysical index of discriminability (Green & Swets, 1966), and here measures the discriminability of a reinforcer on cycle c_i. That index does not generally refer to a sequence of stimuli occurring over a period of time, however (Nevin, 1969, 1981).

Representing Perceived Value

The most elemental form of analysis of the discrimination data examines the choices that an organism makes following a substimulus presentation. The *bias* value represents the tendency to indicate that the substimulus comes from the rich schedule. Why examine bias and not simply discriminability? The reason is that there is no way to figure out discriminability on an individual substimulus. A hit is correctly indicating rich when the rich schedule is presented. But the rich schedule consists of a distribution of all the substimuli, not just one of them. Therefore, there is no way to construct a hit or false alarm rate for individual substimuli. However, there is a very easy way to construct the proportion of left-key responses, p(L), out of the total responses for each substimulus. That is why we are examining bias and not discriminability. There is, however, a relationship between the bias and discriminability at the overall level (Commons, 1981).

A left-key peck to a substimulus is the decision that indicates that the substimulus is rich. Although only a single left- or right-key peck can occur on each dichotomous choice trial, $p(L|S_n)$ indicates the proportion of times the pigeon indicates that the sample was from the rich schedule.[3] The proportion (relative frequency) of a left-key peck in response to the given substimulus is:

$$p(L|S_n) = (L|S_n)/(L|S_n + R|S_n). \tag{4}$$

This proportion is called the bias for a given substimulus sample, S_n, in signal-detection theory (Commons, 1979, 1981). That proportion also indicates the perceived richness of the sample in the following sense. Because the bias proportion is estimated by probability, it can be linearized by a probit transformation, $z^{-1}p(L)$, as shown in Commons (1981). The value, $z^{-1}p(L)$, the inverse-probability (probit) transform of bias, is the *perceived value,* V_{sn}, of the substimulus schedule sample. This probit transformation makes linear a plot of value versus number of reinforcers in a sample (Commons, 1981). Perceived value is just the sum of the decremented values of each reinforcer. A difference exists between this probabilistic estimate and the perceived outcome because sometimes the pigeon would make a choice before having responded to every cycle. For example, on a schedule of 0001, the pigeon might respond after the second or third cycle and never get the reinforcer on the last cycle.

Linear regressions of perceived value were carried out to see what the contributions were of a reinforcer on a cycle. The coefficients, a_i, which

---

[3] This proportion is simplified to p(L).

represent the contributions, were estimated using the following multivariate equation:

$$V_{\mathrm{sn}} = V_{\mathrm{sn}}(L, R) = z^{-1}p(L|S_{\mathrm{n}}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} a_{\mathrm{i}}v_{\mathrm{i}} \,. \tag{5}$$

The value-weights, a_{i}, are applied to the reinforcer-occurrence variable v_{i}. The reinforcer-occurrence variable $v_{\mathrm{i}} = 1$ when a reinforcer did occur and $v_{\mathrm{i}} = 0$ when a reinforcer did not occur on the instance *i* time periods before choice where a reinforcer might be programmed. An a_{i} coefficient reflects the value of a reinforcer at the time of choice when it occurred *i* time periods before the choice. This is because it represents the contribution of the presence of a reinforcer on cycle$_{\mathrm{i}}$ to the prediction that the probability of the choice was 1. The probability of a left choice should be 1 if the stimulus was rich (0 is the value for a lean choice).

Insensitivity-The δ Measure

Here we will introduce a new variable, δ_{i}, which we will call the *inverse of sensitivity* to the signal "reinforcement." It will be represented by the numerical value of $4/(d')^2$ (Egan, 1975; Green & Swets, 1966; Nevin, 1969, 1981). This inverse of sensitivity is introduced so that the form of the following equations is simplified.[4] The general form of the following equations are shown in the appendix and the cases described by the experiment are shown here. The probability density function (PDF) of the amount of reinforcement remembered on a cycle, y_{i}, in the case of the rich schedule, $S_{\mathrm{r}}: v_{\mathrm{i}} = 1$ with probability 3/4 and $v_{\mathrm{i}} = 0$ with probability 1/4, is:

$$f(y_{\mathrm{i}}) = \frac{3}{4}\,\frac{1}{\sqrt{2\pi}\,\sigma_{\mathrm{i}}} \cdot e^{\frac{-(y_{\mathrm{i}}-\mu_{\mathrm{i}})^2}{2\sigma^2_{\mathrm{i}}}} + \frac{1}{4}\,\frac{1}{\sqrt{2\pi}\,\sigma_{\mathrm{i}}} \cdot e^{\frac{-(y_{\mathrm{i}}+\mu_{\mathrm{i}})^2}{2\sigma^2_{\mathrm{i}}}} \,.$$

Here, reinforcement on the rich key comes with probability 3/4, so the normal distribution for reinforcement on the rich key has a coefficient of 3/4, and nonreinforcement only occurs with probability 1/4 so that is the coefficient of the other normal distribution. This probability density function, which is the expectation, is simply the weighted values of

---

[4] When variables are normally distributed, detectability of reinforcement density (Commons, 1979) is $d' = z^{-1}p(\text{Hit}) - z^{-1}p(\text{False Alarm}) = z^{-1}p(\text{Indicates Rich}|\text{Was Rich}) - z^{-1}p$ (Indicates Rich|Was Lean).

reinforcement or nonreinforcement. The expectation is simply the weighted values of getting reinforcement or nonreinforcement. The expected value of the random variable y_i on any given trial, and has the form:

$$\sum_{i=1}^{n} p_i \, v_i$$

where v_i is the value of the random variable on cycle i, p_i is the probability of that value occurring, and n is the number of cycles in the trial.

Assume that probabilistic responding is a result of the combination of signal and random noise in the perceptual and memory processing of reinforcers in the schedule sample. With reasonable amounts of noise, the perceived density will be higher when the actual mean density is higher (Commons, 1979).

As in the case of the rich schedule, in the case of the lean schedule, S_L: $v_i = 1$ with probability 1/4, $v_i = 0$ with probability 3/4, y_i will have density function:

$$f(y_i) = \frac{3}{4}\frac{1}{\sqrt{2\pi\sigma_i}} \bullet e \, \frac{-(y_i-\mu_i)^2}{2\sigma^2_i}$$

$$+ \frac{1}{4}\frac{1}{\sqrt{2\pi\sigma_i}} \bullet e \, \frac{-(y_i+\mu_i)^2}{2\sigma^2_i} \, .$$

If a bird chooses left or right so as to maximize the expected payoff, given the particular remembered values y_1, y_2, y_3, y_4, of the sequence of events v_1, v_2, v_3, v_4 in the scheduled sample, it will employ a likelihood ratio criterion. That is, it will peck left only when the ratio

$$\frac{EP(L|y_1, y_2, y_3, y_4)}{EP(R|y_1, y_2, y_3, y_4)} > 1.$$

Now,

$$\frac{EP(L|y_1, y_2, y_3, y_4)}{EP(R|y_1, y_2, y_3, y_4)} = \frac{U_L \, p(y_1, y_2, y_3, y_4)}{U_R \, p(y_1, y_2, y_3, y_4)} = \frac{U_L \pi_i f(y_i|S_R)}{U_R \pi_i f(y_i|S_R)}$$

Where U_L and U_R represent the utility values for the reinforcers received for a left or right key peck

$$= \frac{U_L}{U_R} \frac{\prod_{i=1}^{4}\left[\frac{3}{4} e \, \frac{-(y_i-\mu_i)^2}{2\sigma_i^2} + \frac{1}{4} e \, \frac{-(y_i+\mu_i)^2}{2\sigma^2}\right.}{\prod_{i=1}^{4}\left.\frac{1}{4} e \, \frac{-(y_i-\mu_i)^2}{2\sigma_i^2} + \frac{3}{4} e \, \frac{-(y_i+\mu_i)^2}{2\sigma_i^2}\right]} ,$$

Taking out the $\Pi(1/4)$ from top and bottom and then expanding, yields

$$= \frac{U_L}{U_R} \prod_{i=1}^{4} \frac{3\, e^{\frac{-(y_i^2 - 2y_i\,\mu_i + \mu_i^2)}{2\sigma_i^2}} + e^{\frac{-(y_i^2 + 2y_i\mu_i + \mu_i^2)}{2\sigma^2}}}{e^{\frac{-(y_i^2 - 2y_i\mu_i + \mu_i^2)}{2\sigma_i^2}} + 3\, e^{\frac{-(y_i^2 + 2y_i\mu_i + \mu_i^2)}{2\sigma_i^2}}},$$

Then the term, $e^{\frac{-(y_i^2+\mu_i^2)}{2\sigma^2}}$ may be factored out and canceled top and bottom, and simultaneously dividing top and bottom by the term

$e^{\frac{-(y_i\mu_i)}{2\sigma_i^2}}$ yields

$$= \frac{U_L}{U_R} \prod_{i=1}^{4} \frac{3\, e^{\frac{(2y_i\mu_i)}{\sigma_i^2}} + 1}{3 + e^{\frac{(2y_i\mu_i)}{\sigma_i^2}}},$$

The bird will peck left when this quantity is greater than 1. We see that this equation is a nonlinear condition on the y_i values. But in the case that the discrimination of the reinforcers is difficult (that is, all the δ_i are relatively large), and therefore $\mu_i << \sigma_i^2$ which would appear to be the case in this experiment, the following truncated Taylor series expansions of the exponential may be used as approximations:

$$\frac{3\, e^{\frac{(2y_i\mu_i)}{\sigma_i^2}} + 1}{3 + e^{\frac{(2y_i\mu_i)}{\sigma_i^2}}} \cong \frac{3\left[1 + \frac{2y_i\mu_i}{\sigma_i^2} + \frac{2{y_i}^2{\mu_i}^2}{{\sigma_i}^4}\right] + 1}{3 + \left[1 + \frac{2y_i\mu_i}{\sigma_i^2} + \frac{2{y_i}^2{\mu_i}^2}{{\sigma_i}^4}\right]},$$

Dividing top and bottom by 4/3

$$\cong \frac{1 + \frac{3}{2}\frac{y_i\mu_i}{{\sigma_i}^2} + \frac{3}{2}\frac{{y_i}^2{\mu_i}^2}{{\sigma_i}^4}}{1 + \frac{1}{2}\frac{y_i\mu_i}{{\sigma_i}^2} + \frac{1}{2}\frac{{y_i}^2{\mu_i}^2}{{\sigma_i}^2}},$$

Multiplying top and bottom by

$$\cong \left[1 + \frac{3}{2}\frac{y_i\mu_i}{\sigma_i^2} + \frac{3}{2}\frac{y_i^2\mu_i^2}{\sigma_i^2}\right]\left[1 - \frac{1}{2}\frac{y_i\mu_i}{\sigma_i^2} - \frac{1}{4}\frac{y_i^2\mu_i^2}{\sigma_i^2}\right]$$

yields

$$\cong 1 + \frac{y_i\mu_i}{\sigma_i^2} + \frac{1}{2}\frac{y_i^2\mu_i^2}{\sigma_i^2}, \tag{5a}$$

and finally

$$\cong e^{y_i\mu_i}/\sigma_i^2.$$

This is the formula for the four-cycle data collected by Commons (1981). A more generalized form of this equation is derived in the appendix, where a sample of *n*-cycles over any symmetrical reinforcement densities is analyzed as above.

So, for four-cycle data, the condition for the bird to peck left is

$$\frac{U_L}{U_R}\prod_{i=1}^{4} e^{\frac{y_i\mu_i}{\sigma_i^2}} > 1.$$

That is,

$$\prod_{i=1}^{4} e^{\frac{y_i\mu_i}{\sigma_i^2}} > \frac{U_R}{U_L}.$$

Taking the natural log of both sides gives

$$\sum_{i=1}^{4}\frac{y_i\mu_i}{\sigma_i^2} > \ln\frac{U_R}{U_L}.$$

Note that both linear and quadratic terms in each of the four Taylor expansions are kept, so that the result is correct up to terms of order $(1/\delta_i)^3$; when δ_i is very large, $(1/\delta_i)^3$ is very small. Furthermore, we only care about the validity of the approximation in the region where $EP(L)/EP(R) \cong 1$; that is, where

$$\sum_{i=1}^{4}(y_i\mu_i)\,/\,\sigma_i^2 \cong \ln(U_R/U_L),$$

which in the case of only moderate bias, will mean y_i values such that none of the $(y_i\mu_i)/\sigma_i^2 = y_i/\delta\mu_i$ is very large. The term on the left has been derived directly from the PDF given.

Here we have shown that the bird will peck left when

$$\sum_{i=1}^{4} \frac{y_i \mu_i}{\sigma_i^2} > 1\text{n}\,\frac{U_R}{U_L}\ ,$$

But the expression on the left, as a sum of normally distributed multivariate random variables, is itself normally distributed,

$$M_i = \sum_{i=1}^{4} \frac{(2\nu_i - 1)}{\delta_i}\quad \text{, which is the mean}$$

$$V_{S\ n} = \sum_{i=1}^{4} \frac{1}{\delta_i}\quad \text{, which is the variance}$$

and is conditional on the occurrence of the substimulus described by the v_i, since the conditional distribution of y_i is $N(\mu_i\,(2\nu_i\ \text{-}1), \sigma_i^2)$. Therefore the probability of its being greater than $1\text{n}\ (U_R/U_L)$, given the occurrence of that substimulus is given by

$$z_{p(L)} = \frac{M - \mu_i}{\sigma_i} = \frac{M - 1\text{n}(U_R/U_L)}{V_{sn}^{\ 1/2}}$$

so that

$$z_{p(L)} = \sum_{i=1}^{4} a_i v_i + b\ ,$$

where

$$a_i = \frac{2}{\delta_i \left[\sum_{j=1}^{4} \frac{1}{\delta_j} \right]^{1/2}} \tag{6}$$

The a_i coefficients can now be interpreted as indicating the rate at which discriminability of reinforcement declines with time. The value of δ_i (indistinguishability) was defined as $4/(d')^2$. The set of δ_i implied by a given set of a_i coefficients are easily reconstructed by inversion of Equation (6):

$$\sum_{i=1}^{4} a_i = \sum_{i=1}^{4} \frac{2}{\delta_i \left[\sum_{j=1}^{4} \frac{1}{\delta_j} \right]^{1/2}}$$

$$= \left[\sum_{i=1}^{4} \frac{1}{\delta_i}\right]\left[\frac{2}{\left[\sum_{j=1}^{4} \frac{1}{\delta_j}\right]^{1/2}}\right]$$

$$= 2\left[\sum_{i=1}^{4} \frac{1}{\delta_i}\right]^{1/2} .$$

From Equation 6 we get

$$\delta_i = \frac{2}{a_i\left[\sum_{j=1}^{4} \frac{1}{\delta_i}\right]^{1/2}}$$

$$= \frac{4}{a_i\left[\sum_{j=1}^{4} a_j\right]}$$

The mean μ_i is the value of reinforcement remembered (i.e., perceived value) for a given cycle. Since the contributions of reinforcers will be added, the detectabilities should be in a form in which they add easily. By having the variance in the numerators, it will be possible to add the detectabilities, since the variance of the sum of independent variables equals the sum of the individual variances. The inverse of discriminability of reinforcement in cycle$_i$ is then measured by:

$$\delta_i = \frac{\sigma_i^2}{\mu_i^2} . \tag{7}$$

The parameter δ_0 is the inverse of sensitivity of reinforcement that occurs with no delay between reinforcement and choice. Note that when $\delta_i = 0$, variance is 0 and detection is perfect. Likewise, as δ_i gets larger, detection becomes poorer.

The simplest model of forgetting would be to assume that the *signal* stays constant while the amount of *noise* increases linearly with time. If it is assumed that the perceived value Y at the time of reinforcement is a random variable Y_0 (random because of imperfect discrimination of reinforcement even at that time), and that with each cycle that passes thereafter there is added another random noise component e (with the random noise terms independent of Y_0 and of each other, and all identically distributed), then the random variable Y_i, representing remembered values of reinforcement after i cycles have passed (in the case of the four-cycle experiment), can be written as

$$Y_i = Y_0 + \sum_{j=1}^{4} e_j \, .$$

Therefore,

$$E(Y_i) = E(Y_0) \, ,$$

$$\text{var}\,(Y_i) = \text{var}\,(Y_0) + i * \text{var}\,(e) \, ,$$

and

$$\delta_i = \frac{\text{var}\,(Y_i)}{E(Y_i)^2} = \frac{\text{var}\,(Y_0)}{E(Y_0)^2} + \text{i} * \frac{\text{var}\,(e)}{E(Y_0)^2} \, .$$

The discriminability measure δ_i, as a function of cycle number i, is therefore predicted to be a linear function. The first term in the last equation is replaced by its equivalent, δ_0; the second term by its equivalent, $i\,\gamma$:

$$\delta_i = \delta_0 + i\,\gamma \, .$$

As noise is linearly added to the initial inverse of discriminability, the variance goes up so δ_i increases. Hence, according to a simple additive model, the inverse of discriminability of reinforcement on cycle$_i$ is simply a function of the initial discriminability plus the product of τ, the rate at which noise is added between the reinforcer and choice, and γ, the rate of forgetting:

$$\delta_i = \delta_0 + \tau\,\gamma, \text{ for some } \delta_0 \text{ and } \gamma > 0. \tag{8}$$

How the General Additive Noise Model Predicts Perceived Reinforcement Value

The rate at which noise is added between the reinforcer and choice, τ, can be represented by the product of three terms: i, t^{α}, and s^{β}. The parameter i represents how far (in time periods) a possible reinforcer is away from choice. The value t is the length of the standard cycle and s is the probe multiplier ($s = 1$ for standard, $s = 2$ for double, $s = 3$ for triple). Note that the passage of time or elapsed time equals $i\,t\,s$. Three submodels, which are created by setting α, β, or both equal to 1 or 0, generate three notions of the rate of noise addition. If there is no effect of cycle length on perceived value, $\alpha = 0$. If there is no effect of probe multiplier on perceived value, $\beta = 0$. These submodels will be presented.

Each of the submodels is subsumed by the more general model:

$$\delta_i = \delta_0 + i\, t^{\alpha} s^{\beta} \gamma\,. \tag{9}$$

This equation might be used to estimate the four parameters: δ_i, α, β, γ by employing the data for all substimuli, all cycle lengths, and all probes, and special hypotheses (e.g., submodel 1 assumes $\alpha = \beta = 1$) subjected to a statistical hypothesis test (chi-square).

To show that value decreases as a hyperbolic function of i, the representation of value in the present experimental example is defined. As explained in Commons, Woodford, and Ducheny (1982) and here, the perceived reinforcement density, $V_{s\,n}$ of a sample, S_n, is a function of the δ's. That value is defined as the bias to indicate rich, $z^{-1}p(L)$. Again, it is assumed that the value of the entire substimulus is just the value of the sum of the contributions, $a_i v_i$, which is the effect of each reinforcer as remembered: Then the value of a sample as a function of δ_i is derived in the appendix. That derivation shows that the value of a sample should fit a hyperbolic model. The results are as follows:

$$V_{s_n} = z^{-1}p(L) = \sum_{i=1}^{4} a_i v_i\,, \tag{2, 10}$$

where

$$a_i = [\delta_i\, (\sum_{j=1}^{4} \frac{2}{\delta_j})^{1/2}\,]^{-1}\,. \tag{11}$$

The value V_{s_n} in hyperbolic form is as shown:

$$z^{-1}p(L) = \sum_{i=1}^{4} \frac{c}{d + i}\, v_i\,, \tag{12}$$

where

$$c = \frac{\left[t^{\alpha} s^{\beta} \gamma \right]^{-1}}{\left[\sum_{j=1}^{4} \frac{1}{\delta + j t^{\alpha} s^{\beta} \gamma} \right]^{1/2}}\,, \tag{13}$$

$$d = \frac{\delta_0}{t^{\alpha} s^{\beta} \gamma}\,. \tag{14}$$

Equation 12 shows that the equation for reinforcer value has a hyperbolic form. The parameter *i* represents how far (in cycles) a possible reinforcer is away from choice. Equations 12, 13 and 14 give the predicted effects of varying *t* and *s* on the parameters of the hyperbolic model. There are several plausible hypotheses about how the rate of forgetting (γ) varies as the length of the cycles varies. The submodels are presented next.

Submodel 1 (Time)

It might be supposed that the rate at which noise is added to the memory of a reinforcer depends solely on the passage of time, rather than on intervening events or a combination of events and time. In this case sensitivity on cycle$_i$ would be

$$\delta_i = \delta_0 + i\,t\,s\,\gamma. \tag{15}$$

The value t = length of the standard cycle and s = probe multiplier (s = 1 for standard, $s = 2$ for double, $s = 3$ for triple). Again, elapsed time equals $i\,t\,s$. In Equation 15, δ_0 and γ should be the same for all cycle lengths and probes. Neither δ_0 nor γ is affected by time.

To see if submodel 1 fits the data, a linearization of bias measured by $z^{-1}p(L)$, which is the perceived value of the schedule sample, is examined as a function of time in Fig. 7.3.

If memory were perfect, then perceived value of the sample for all substimuli taken together would be zero, $z^{-1}p(L) = 0$. Any indiscriminate change from the average bias would either increase false alarms (indicating rich when it was in fact lean) or misses (indicating lean when it was rich). Decreases in $z^{-1}p(L)$, therefore, represent forgetting. As standard cycle length t increased, $i\,t\,s$ would increase and submodel 1 would predict a decrease in discriminability. This would appear as a decrease in $z^{-1}p(L)$. This decrease was not found in the data shown in the left column of panels in Fig. 7.3. Submodel 1 would predict that the substimuli with the shorter cycles (circles) should be better discriminated than the middle length cycles (triangles) and those in turn better than the long cycle length (squares). The base-length-cycle substimuli were all equally well discriminated, which does not support submodel 1.

Submodel 2 (Number of Intervening Cycles)

The contrary extreme hypothesis would be to suppose that the rate at which noise is added depends solely on the number of subsequently occurring events between the reinforcer in question and choice, and not on the amount of time that elapsed. The number of cycles (number of events

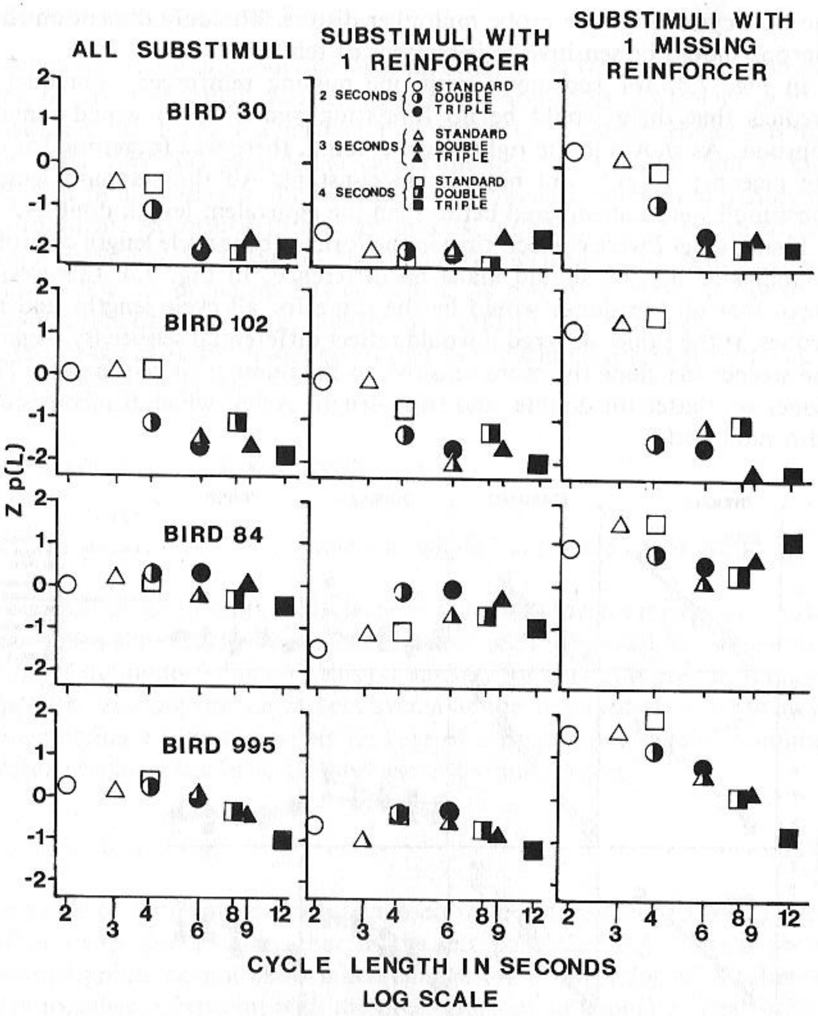

Fig. 7.3 Perceived density, $z^{-1}p(L)$, plotted against cycle length in log seconds. The left-hand column shows the effect of cycle length on the perceived density of all substimuli considered together. This is the macrolevel relation. At the molar level, Columns 2 and 3 show the relation between perceived density and cycle length for substimuli with one reinforcer and substimuli with three reinforcers (one missing reinforcer).

to store) is all that matters. Here those intervening events are the cycles with possible reinforcers in them. In that case it would be expected that:

$$\delta_i = \delta_0 + i\,\gamma. \qquad (16)$$

In this submodel, δ_0 and γ values should be the same for all cycle lengths and probes. This means each a_i is identical to the corresponding a_i although the base-cycle length or probe multiplier differs. The only dimension that pigeons should be sensitive to is number of reinforcers.

In Fig. 7.3, for substimuli with one missing reinforcer, submodel 2 predicts that there would be no forgetting and $z^{-1}p(L)$ would remain constant. As shown in the right -most column, there was forgetting for all the pigeons; $z^{-1}p(L)$ did not remain constant. All the standard length substimuli were remembered better than the equivalent length doubles.

If submodel 2 were correct, time in the form of base cycle length or probe doubling or tripling should make no difference. In Fig. 7.4, this would mean that all the slopes would be the same for all cycle lengths and all probes. If the slopes differed it would reflect differential sensitivity. Again, the steeper the slope the more sensitive to the number of reinforcers. The slopes get flatter for double- and triple-length cycles, which is inconsistent with submodel 2.

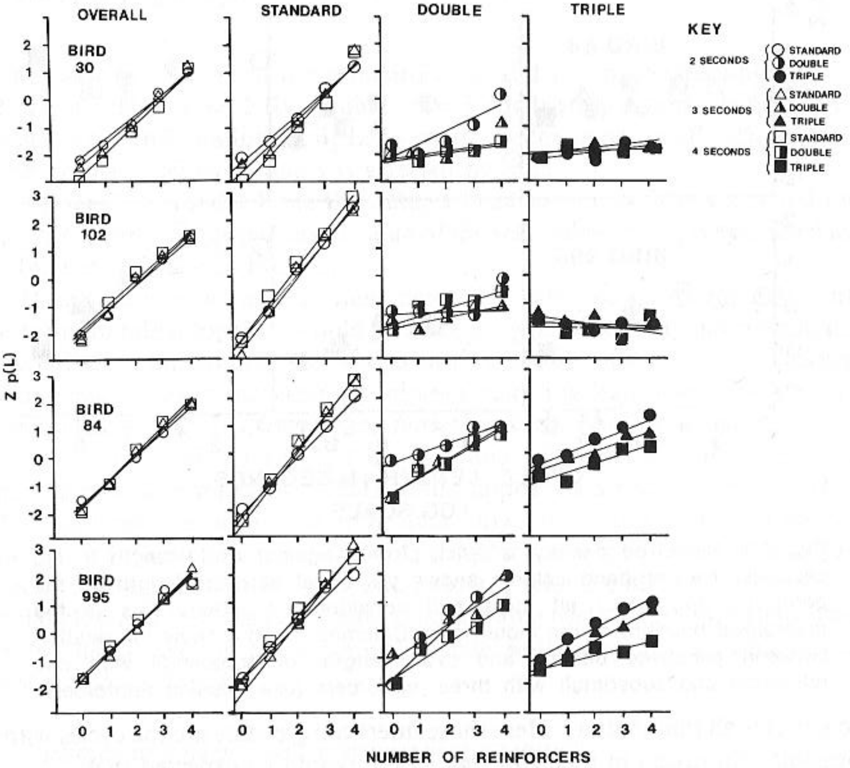

Fig. 7.4 Perceived density, $z^{-1}p(L)$, shown as a function of actual density (the number of reinforcers in a substimulus). This molar relation is shown for all trials together (Column 1), for trials with standard cycle length (Column 2), for probe trials where the standard length was doubled (Column 3) and tripled (Column 4).

Submodel 3 (Relative Time)

An intermediate submodel assumes that the rate at which the remembered value of a reinforcement decreases is adjusted according to the average rate at which events occur in a particular experimental situation. Hence it is not how much time has passed that causes the decrease in value, but simply length of time relative to the average time between events in the particular environment. According to this hypothesis, a change in the length of the standard cycle would have no effect, but the inclusion of a probe substimulus when the organism has calibrated its rate of memory decay to a different rate of occurrence of events will affect the size of the a_i. This is a finite storage capacity model given that probe stimuli are interpreted as a larger number of events:

```
x - - x - - - x x     events

. . . . . . . . .     points at which reinforcers could occur.
```

The pigeon expects events at all the points and so allocates memory space to each. This submodel is especially consistent with submodel 3a, a generalization of submodel 3 that contains parameters for bias. It is presented in the appendix. The pigeon remembers events at the standard-cycle frequency. Remembering all the nonevents (in case of a probe) as events of nonreinforcement causes the bias. In this case one would expect:

$$\delta_i = \delta_0 + i\, s\, \gamma. \tag{17}$$

The value of $i\,s$ increases with increased probe multipliers or with cycles further from choice. Note that in the left panels of Fig. 7.3 the three left-most points do not show much change for $z^{-1}p(L)$ for all substimuli taken together, consistent with the predictions of submodel 3. That is, the perceived value of reinforcement did not change for the 2, 3, and 4 second base cycle lengths. The very problems that submodel 1 and submodel 2 had are solved by submodel 3. The effects of both probes and time are predicted. For instance, the 4-s standard, shown in the right column, is always above the 2-s double probe.

Submodel 3 was generally supported by the data shown in Fig. 7.4. For this submodel, if the probe multiplier s increased, all the δ_i's should increase. Increased δ_i's would indicate decreased sensitivity—the slope of V_{sn} versus number of reinforcers should decrease. In Fig. 7.4, for all pigeons, the slopes for the standard base cycle length are similar. As the probe cycle-length multiplier increased, the slope decreased as submodel 3 predicts. The data points for trials with standard cycle lengths fell on top of each other for all four birds. These points began to diverge for trials of double standard length (i.e., as *s* increased, base trial length made a greater difference in slope), and diverged even more for trials of triple standard length in the cases of Pigeons 84 and 995. This is all consistent with submodel 3; an increase in *s* should cause a greater divergence in slope as base cycle length increases because δ_0 increases as base cycle length increases, and therefore δ_i does as well. But this increase is linear, the sum of increased $s\ \gamma$ and δ_0, and thus divergence should be linear as well. For pigeons 84 and 995, this assumption is supported.

The exceptions to submodel 3 can be seen in the *Triple* column of Fig. 7.4. Pigeons 30 and 102 show extreme bias: Their responses to probes are severely biased in favor of the lean reinforcer. Note that none of these submodels predicts the value of a bias term for the entire hyperbolic model. Hence, birds that react to the probe trials in a severely biased fashion cannot be used to distinguish any of these hypotheses. An alternative explanation exists, however, for the data that indicate bias as shown for the pigeon 30 and 102 in the top middle panels of Fig. 7.4.

Fully General Model with Parameters for Bias, Models 1a, 2a, 3a

Pigeons 30 and 102 exhibited extreme bias for going to the right for triples as shown in Fig. 7.4. This is consistent with submodel 3a, a generalization of submodel 3 that adds terms to estimate this form of bias. A fully general model with the parameter *s* for this form of right bias is shown in the appendix. If this bias parameter, *s*, is small in submodel 3a, then the tendency to choose left would increase. If it is exactly 1, then submodel 3a reduces to just submodel 3. If *s* is very large then the value of a substimulus is markedly decreased, resulting in a strong right bias. This is just what was found as shown in the top middle panels of Fig 7.4. There is no general theoretical basis at this time for expecting such right bias, other than the possibility that all events outside a time period fixed by the length of the cycles ($2 \times 4 = 8$ s, $3 \times 4 = 12$ s, $4 \times 4 = 16$ s) are totally ignored. If these events are ignored, then subjects act as if they had zero value. Another possibility is that longer probe cycles result in memory of a large number of nonreinforcements (without truncation of time period over which memory occurs).

CONCLUSION

It was shown analytically that the time window must be a hyperbolic function of value versus some measure of time. The major assumption made was that noise interfering with the effectiveness of reinforcement is added in a linear fashion.

Applying the general additive noise model to the results from Commons (1981) showed that the decrementation function is not a simple function of simple time. The effectiveness of reinforcement decreased as a function of a form of time. This form of time progressed along the relative rate dimension; thus time progressed relative to the rate of baseline reinforcement. Three submodels and a set of variants were qualitatively tested. The experimental requirement was for the pigeons to detect the richness of the sample. Therefore, the number of reinforcers in the sample determined which choice would be most likely to be reinforced. The best decision rule required sensitivity only to number of reinforcers. Out of the possible four reinforcers, if there were zero or one reinforcers the best choice would be to indicate lean; if there were three or four reinforcers the best choice would be rich. Other decision rules could mean that the pigeons are less sensitive in some way since they would not exhibit an exclusive preference. Mazur (1984, 1987) leaves time in one schedule constant while varying delay in another. Until background rate of reinforcement is varied, the role of context per se in the choice situations consisting of a single or a few reinforcers will not be clear. From the perspective of melioration theory (Herrnstein, 1982; Herrnstein & Vaughan, 1980) and our own acquisition theory (Commons, Woodford, Boitano, Ducheny, & Peck, 1982) the background rate of reinforcement as well as its variability should be important as we have found here.

Qualitative fits of submodels of the general additive noise model showed that neither simple time nor the number of reinforcements (i.e., the number of cycles with reinforcement) accounted for the findings. A test of the third submodel suggested that a parameter for overall bias independent of experimental conditions should be included. The decrease in reinforcement value was sensitive to changes in reinforcement rate with respect to background rate and not with respect to number of reinforcements or time. The general additive noise model can be used in preference studies since the model was formulated in terms of "time" between a response and a reinforcing schedule component. The schedules modeled here are general in form, since they are T-schedules that resemble random interval schedules. As Schoenfeld and Cole (1972) have shown, T-schedules with certain parameter selections behave like interval schedules. From the assumption that noise is linearly added between reinforcement and choice, hyperbolic decrementation functions were analytically derived and then illustrated.

APPENDIX

The value of a substimulus, V_{sn}, is the bias to indicate that a substimulus is rich, $z^{-1}p(L)$. That value is directly related to δ_i. Again, it is assumed that the value of the entire substimulus is just the value of the sum of the contributions, $a_i v_i$, which is the effect of each reinforcer as remembered. This derivation gives the predicted effects of varying t and s on the parameters of the hyperbolic model. The second important result of the derivation is that our use of the additive noise model yields a value of a sample that fits a hyperbolic model.

$$V_{s_n} = z^{-1}p(L) = \sum_{i=1}^{4} a_i v_i \ , \tag{2, 10}$$

where

$$a_i = \frac{2}{\delta_i \left[\sum_{j=1}^{4} \frac{1}{\delta_j} \right]^{1/2}} \ . \tag{11}$$

Fully General Submodel: Submodel 3a

By Equation 13, which states that $\delta_i = \delta_0 + i\, t^\alpha s^\beta \gamma$, and Equation 2, 10, and 11,

$$v_{s_n} = 2 \sum_{i=1}^{4} \ [\ (\delta_0 + i\, t^\alpha s^\beta \gamma)\ [\sum_{j=1}^{4} \frac{1}{\delta_j}]^{1/2}]^{-1} v_i \tag{18}$$

$$= 2 \sum_{i=1}^{4} \ [\ (t^\alpha s^\beta \gamma) \frac{(\delta_0 + i)}{t^\alpha s^\beta \gamma}\ [\sum_{j=1}^{4} \frac{1}{\delta_i}\ ]^{1/2}]^{-1} v_i$$

$$= \frac{2 \sum_{i=1}^{4} \ [\ (t^\alpha s^\beta \gamma)\ [\sum_{j=1}^{4} \frac{1}{\delta_j}\ ]\ ]^{1/2} v_i^{-1}}{\frac{\delta_0}{t^\alpha s^\beta \gamma} + i} \ . \tag{20}$$

Equation 21, which is derived from Equation 20, shows that the decrementation function is hyperbolic in form:

$$z^{-1}p(L) = \sum_{i=1}^{4} \frac{c}{d + i} v_i \ , \tag{12, 21}$$

where

$$c = [\, (t^{\alpha} s^{\beta} \gamma)\, [\sum_{j=1}^{4} \frac{2}{\delta_j}\,]^{1/2}\,]^{-1}\,, \qquad (13,22)$$

$$d = \frac{\delta_0}{t^{\alpha} s^{\beta} \gamma}\,. \qquad (14,23)$$

Generalizing to *n* Cycles and any Symmetrical Reinforcement Densities

The models in the text are generated for four-cycle data only. Fully generalized derivations of the Probability Density Function (PDF) and δ_i for n cycles are as follows:

For the rich schedule where a left response indicates reinforcement the probability density function has the form

$$f(y_i) = p(R|S_R)\frac{1}{2\mathit{II}\,\sigma_i} \bullet e\,\frac{-(yi - \mu_i)^2}{2\sigma_i^2} + p(N|S_R)\,\frac{1}{2\mathit{II}\,\sigma_i} \bullet e\,\frac{(y_i + \mu_i)^2}{2\sigma_i^2}\,.$$

Likewise, for the lean schedule where reinforcement is delivered on the right response, the probability density function has the form

$$f(y_i) = p(R|S_L)\frac{1}{2\mathit{II}\,\sigma_i} \bullet e\,\frac{-(yi - \mu_i)^2}{2\sigma_i^2} + p(N|S_L)\,\frac{1}{2\mathit{II}\,\sigma_i} \bullet e\,\frac{-(y_i + \mu_i)^2}{2\sigma_i^2}.$$

Where $p(R|S_n)$ is the probability of reinforcement on the given substimulus, (i.e., $v_i = 1$) and $p(N|S_n)$ the probability of no reinforcement on the substimulus. The condition for the bird to peck left

$$\frac{EP(L|y_1, y_2, y_3, y_4 \ldots y_n)}{EP(R|y_1, y_2, y_3, y_4 \ldots y_n)} > 1\,,$$

$$\frac{EP(L|y_1, y_2, y_3, y_4 \ldots y_n)}{EP(R|y_1, y_2, y_3, y_4 \ldots y_n)} = \frac{U_L}{U_R}\,\frac{p(y_1, y_2, y_3, y_4 \ldots y_n)}{p(y_1, y_2, y_3, y_4, \ldots y_n)}$$

$$= \frac{U_L \mathit{II}_i f(y_i|S_R)}{U_R \mathit{II}_i f(y_i|S_R)}\,,$$

$$= \frac{U_L}{U_R} \frac{\prod_{i=1}^{n} \left(p(R|S_R)\, e^{\frac{-(y_i-\mu_i)^2}{2\sigma_i^2}} + p(N|S_R)\, e^{\frac{-(y_i+\mu_i)^2}{2\sigma_i^2}}\right)}{\prod_{i=1}^{n} \left(p(R|S_L)\, e^{\frac{-(y_i-\mu_i)^2}{2\sigma_i^2}} + p(N|S_L)\, e^{\frac{-(y_i+\mu_i)^2}{2\sigma_i^2}}\right)}.$$

Factoring and expanding, as shown, yields

$$= \frac{U_L}{U_R} \prod_{i=1}^{n} \frac{p(R|S_R)\, e^{\frac{-(y_i^2-2y_i\mu_i+\mu_i^2)}{2\sigma_i^2}} + p(N|S_R)\, e^{\frac{-(y_i^2+2y_i\mu_i+\mu_i^2)}{2\sigma_i^2}}}{p(R|S_L)\, e^{\frac{-(y_i^2-2y_i\mu_i+\mu_i^2)}{2\sigma_i^2}} + p\,(N|S_L)\, e^{\frac{-(y_i^2+2y_i\mu_i+\mu_i^2)}{2\sigma_i^2}}}.$$

Factoring and dividing again, as shown, yields

$$= \frac{U_L}{U_R} \prod_{i=1}^{n} \frac{p(R|S_R)\, e^{\frac{(2y_i\mu_i)}{\sigma_i^2}}\, p(N|S_R)}{p(N|S_L) + p(R|S_L)\, e^{\frac{(2y_i\mu_i)}{\sigma_i^2}}}.$$

For the next part of this discussion, it is sufficient to work with only the function within the iterative multiplication. The function in this product is approximated by the Taylor expansion

$$\cong \frac{p(R|S_R)\left[1 + \frac{2y_i\mu_i}{\sigma_i^2} + \frac{2y_i^2\mu_i^2}{\sigma_i^4}\right] + p(N|S_R)}{p(N|S_L) + p(R|S_L)\left[1 + \frac{2y_i\mu_i}{\sigma_i^2} + \frac{2y^2\mu_i^2}{\sigma_i^4}\right]},$$

and, since this case is symmetrical, expanded to

$$\frac{1 + \frac{p(R|S_R)2y_i\mu_i}{\sigma_i^2} + \frac{p(R|S_R)2y_i^2\mu_i^2}{\sigma_i^4}}{1 + \frac{p(N|S_L)2y_i\mu_i}{\sigma_i^2} + \frac{p(N|S_L)2y_i^2\mu_i^2}{\sigma_i^4}}.$$

Multiply top and bottom by

$$\left[1 - \frac{p(R|S_L)2y_i\mu_i}{\sigma_i^2} - \frac{p(R|S_L)\,(1 - p(R|S_L)\,)2y_i^2\mu_i^2}{\sigma_i^2}\right]$$

and drop all but linear and quadratic terms. The denominator thereby approximates 1 and we are left with the product of the numerator and our polynomial:

$$\left[1 + \frac{p(R|S_R)2y_i\mu_i}{\sigma_i^2} + \frac{p(R|S_R)2y_i^2\mu_i^2}{\sigma_i^4}\right]$$

$$\left[1 - \frac{p(R|S_L)2y_i\mu_i}{\sigma_i^2} - \frac{p(R|S_L)\ (1 - 2p(R|S_L)\ )2y_i^2\mu_i^2}{\sigma_i^4}\right]$$

ACKNOWLEDGMENTS

This work has a long history that would have been terminated except for the inspiration and encouragement of Edmund Fantino. On two occasions Fantino and Ben Williams, along with their students, listened to presentations of our ideas in their formative stages. Parts of this chapter were also reported at the Symposium on Quantitative Analyses of Behavior at Harvard in a paper entitled "The Effects of Delay and Intervening Events" (1982). That paper's authors were Michael L. Commons and Michael Woodford. The data appeared in Commons (1981). Data acquisition was carried out with the support of Richard J. Herrnstein, and with the assistance of Brian Cabral, Wilson R. Fong, and Joseph P. McNamara. Fong assisted with the data analyses. The computer analysis was performed at the Northeastern University academic computer center. The first author served as lecturer in Foundation of Education and in Psychology. William Minty of the Harvard University Drafting Department and Rebecca M. Young drew the graphs.

The extensive editorial help was provided by Michael Davison and John Anthony Nevin, the co-editors. In addition, Leonard S. Miller made extensive substantive suggestions for explicating the theory. William Vaughan, Jr., and Charlotte Mandell commented on the manuscript. Brian Cabral wrote some of the introductory material. Rebecca M. Young, Patrice M. Miller, L. Brockman, and Stanley J. Morse edited the manuscript for style. The research and analysis were supported by grants from the Dare Association, Inc., and NIHM 15494 to Harvard University, Richard J. Herrnstein, principal investigator, and from the Dare Association, Inc., to the first author; and a MacArthur Foundation Award and a National Science Foundation Fellowship to the second author.

REFERENCES

Ainslie, G. W. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82,* 463–496.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation, 8,* 47–90.

Barnes, J.M., & Underwood, B. J. (1959). "Fate" in first-list associations in transfer theory. *Journal of Experimental Psychology, 58,* 97–105.

Chung, S., & Herrnstein, R. J. (1967). Choice and delay of reinforcement. *Journal of the Experimental Analysis of Behavior, 32,* 101–120.

Commons, M. L. (1979). Decision rules and signal detectability in a reinforcement density discrimination. *Journal of the Experimental Analysis of Behavior, 32,* 67–74.

Commons, M. L. (1981). How reinforcement density is discriminated and scaled. In M. L. Commons & J. A. Nevin (Eds.), *Quantitative analyses of behavior: Vol. 1. Discriminative properties of reinforcement schedules* (pp. 51–85). Cambridge, MA: Ballinger.

Commons, M. L., & Nevin, J. A. (Eds.). (1981). *Quantitative analyses of behavior: Vol. 1. Discriminative properties of reinforcement schedules.* Cambridge, MA: Ballinger.

Commons, M. L., Woodford, M., Boitano, G. A., Ducheny, J. R., & Peck, J. R. (1982). Acquisition of preference during shifts between terminal links in concurrent chain schedules. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behavior: Vol. 3. Acquisition* (pp. 391–426). Cambridge, MA: Ballinger.

Commons, M. L., Woodford, M., & Ducheny, J. R. (1982). How reinforcers are aggregated in reinforcement-density discrimination and preference experiments. In M. L. Commons, R. J. Herrnstein, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 2. Matching and maximizing accounts* (pp. 25–78). Cambridge, MA: Ballinger.

Davison, M. C., & Tustin, R. D. (1978). The relation between generalized matching law and signal detection theory. *Journal of the Experimental Analysis of Behavior, 29,* 331–336.

Dickinson, A. (1980). *Contemporary animal learning theory.* Cambridge, England: Cambridge University Press.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans., 1913). New York: Teachers College, Columbia University.

Egan, J. P. (1975). *Signal detection theory and ROC analysis.* New York: Academic Press.

Fantino, E. J. (1981). Contiguity, response strength, and the delay-reduction hypothesis. In P. Harzem & M. H. Zeiler (Eds.), *Advances in analysis of behavior: Vol 2. Predictability, correlation, and contiguity* (pp. 169–201). Chichester, England: Wiley.

Fantino, E. J., Abarca, N., & Dunn, R. M. (1987). The delay-reduction hypothesis: Extensions to foraging and three-alternative choice. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 5. The effect of delay and of intervening events on reinforcement value.* Hillsdale. NJ: Lawrence Erlbaum Associates.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 351–360.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and ROC analysis.* New York: Academic Press.

Grossberg, S. (1978). Theory of human memory: Self organization and performance of sensory motor codes, maps and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology, Vol. 5.* (pp. 233-374). New York: Academic Press.

Hall, G., & Pearce, J. M. (1982). Changes in stimulus associability during conditioning: Implications for theories of acquisition. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behavior: Vol. 3. Acquisition* (pp. 221–240). Cambridge MA: Ballinger.

Herrnstein, R. J. (1982). Melioration as behavioral dynamism. In M. L. Commons, R. J. Herrnstein, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 2. Matching and maximizing accounts* (pp. 433–458). Cambridge, MA: Ballinger.

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.

Madigan, S. A. (1971). Modality and recall order interactions in short-term memory for serial order. *Journal of Experimental Psychology, 87,* 294–296.

Mandell, C. (1981). A psychophysical analysis of time-based schedules of reinforcement. In M. L. Commons & J. A. Nevin (Eds.), *Quantitative analyses of behavior: Vol. 1. Discriminative properties of reinforcement schedules* (pp. 31–50). Cambridge, MA: Ballinger.

Mandell, C. (1984). Discriminability of frequency of food or stimulus presentations in variable-time schedules. *Journal of the Experimental Analysis of Behavior, 42,* 291–303.

Mazur, J. E. (1984). Tests of an equivalence rule for fixed and variable reinforcer delays. *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 426–436.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 5. The effect of delay and of intervening events on reinforcement value.* Hillsdale, NJ: Lawrence Erlbaum Associates.

McCarthy, D.C., & White, K. G. (1987). Behavioral models of delayed detection and their application to the study of memory. *Quantitative analyses of behavior: Vol. 5. The effect of delay and of intervening events on reinforcement value.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Melton, A. W., & Irwin, J. M. (1940). The influence and degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology, 53,* 173–203.

Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64,* 482–488.

Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association, 55,* 299–306.

Nevin, J. A. (1969). Signal detection theory and operant behavior [Review of D. M. Green & J. A. Swets, *Signal detection theory and psychophysics*]. *Journal of the Experimental Analysis of Behavior, 12,* 475–480.

Nevin, J. A. (1981). Psychophysics and reinforcement schedules: An integration. In M. L. Commons & J. A. Nevin (Eds.), *Quantitative analyses of behavior: Vol. 1. Discriminative properties of reinforcement schedules* (pp. 3–27). Cambridge, MA: Ballinger.

Pearce, J. M., Kaye, H., & Hall, G. (1982). Predictive accuracy and stimulus associability: Development of a model for Pavlovian learning. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behavior: Vol. 3. Acquistion* (pp. 241–256). Cambridge, MA: Ballinger.

Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty.* Reading, MA: Addison-Wesley.

Raiffa, H., & Schlaifer, R. O. (1961). *Applied statistical decision theory.* Boston: Division of Research, Harvard Business School.

Schoenfeld, W. N., & Cole, B. K. (1972). *Stimulus schedules: The t-tau systems.* New York: Harper & Row.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652–654.

Underwood, B. J. (1948a). Retroactive and proactive inhibition after five and forty-eight hours. *Journal of Experimental Psychology, 38,* 29–38.

Underwood, B. J. (1948b). "Spontaneous recovery" of verbal associations. *Journal of Experimental Psychology, 38,* 429–439.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64,* 49–60.

White, K. G., & McKenzie, J. (1982). Delayed stimulus control: Recall for single and relational stimuli. *Journal of the Experimental Analysis of Behavior, 38,* 305–312.

Wickelgren, W. A. (1974). Strength/resistance theory of the dynamics of memory storage. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Supper (Eds.), *Contemporary* *developments in mathematical psychology: Learning, memory and thinking*. San Francisco: Freeman.

Williams, B. (1982). Blocking the response-reinforcer association. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behavior: Vol. 3. Acquisition* (pp. 427–448). Cambridge, MA: Ballinger.

Woodword, A. E. (1970). Continuity between serial memory and serial learning. *Journal of Experimental Psychology, 85*, 90–94.