2 HOW REINFORCERS ARE AGGREGATED IN REINFORCEMENT-DENSITY DISCRIMINATION AND PREFERENCE EXPERIMENTS

Michael L. Commons, Michael Woodford, and John R. Ducheny

In the controversy over matching versus maximizing there are two traditions. According to the first, free operant concurrent choice procedures are used; what is thought to be matched or maximized is the amount of reinforcement obtained for responding over a period. Previous illustrations of matching (e.g., Herrnstein, 1970, Chapter 19 of this volume, and Herrnstein and Vaughan, 1980) and maximizing (Rachlin, Kagel and Battalio, 1981), which are accounts of individual choice behavior in operant conditioning experiments, have looked only at the aggregate amount of reinforcement received on a particular schedule. In many situations, such as the concurrent variable-interval, variable-ratio (conc VI VR) schedules described by Herrnstein and Heyman (1979), aggregate data are sufficient to indicate that the experimental subjects do not allocate behavior in a way that would maximize obtained aggregate reinforcement. As Herrnstein has mentioned (personal communication 1981), it is always possible to add constraints that "save" maximizing in this situation. One of these could be that the subjects were not fully sensitive to the contingencies applied to their behavior. Then, the deviation from obtained aggregate maximization would often be roughly consistent with a hypothesis that aggregate relative rate or probability of behavior is "matched" to aggregate obtained relative rate or probability of reinforcement.

Portions of this chapter were reported at the Harvard Symposium on Quantitative Analyses of Behavior, 1979, and the Midwestern Association for Behavior Analysis, Chicago, 1976, by A.A. Thompson, M.P. Krupa, J.J. Andersen, and M.L. Commons, and at the Eastern Psychological Association, Boston, 1977, by M.P. Krupa, M. Hirvonen, J. Pohl, M. Lamb, and A.A. Thompson. These authors collected the 4-cycle discrimination data, wrote the computer programs, and prepared the preliminary reports. The line of research

In the second tradition discrete-trial noncorrection choice procedures are used; what is thought to be matched or maximized is the amount of reinforcement programmed (e.g., Estes, 1957; Neimark and Estes, 1967). However, in a discrete-trial choice situation, the four accounts only make two distinguishable predictions. Matching of obtained reinforcement and maximizing of either obtained or programmed reinforcement all predict that one alternative comes to be chosen to the exclusion of others. The fourth possibility, that of matching of programmed or expected payoff, appears as undermatching of obtained payoff. Estes (1957) proposed a matching law for such expected payoff in a probability learning situation. With pigeons, matching to expected payoff has been found in a discrete-choice reinforcement density discrimination (Commons, 1979, 1981). Here, unless otherwise specified, matching or maximizing refer to programmed reinforcement or expected payoff rather than obtained reinforcement.

The aggregate data in both traditions, however, are ill suited to examination of the possibly more sophisticated local maximizing model proposed here. One of the purposes then of this chapter is to account for undermatching of obtained reinforcement and matching of expected payoff with such a model. It is hypothesized that at the time of each individual choice, the subject acts so as to maximize the expected utility of reinforcement resulting from that particular choice. However it has an incomplete awareness of the effects of the choice that are available to it.[a] More specifically, this model supposes imperfect discrimination of stimuli, including reinforcers, and even more imperfect memory for past

---

leading up to the discrimination experiment was developed under the guidance of John Anthony Nevin. John R. Ducheny collected the 6-cycle discrimination data and he and Joel R. Peck supervised the data collection for the preference experiment. The data acquisition on the preference experiment was carried out with the support of Richard J. Herrnstein. The computer programming was done by Rose P. Meegan and Gilbert Yee with assistance from Wei-o Huang and Katrine Norris. The computer analysis was performed at the Northeastern University Computer Center. The graphs were made by Rose P. Meegan, Brian D. Cabral, Martin N. Davidson, and Rodney Wong. Drafts were read and edited by Patrice M. Miller, Sherry Luna, James E. Mazur, Margaret Vaughan, Theodore L. Allen, and Philip F. Kelleher. The recommendations made by R. Duncan Luce and Richard J. Herrnstein after hearing presentations of the material were invaluable. Howard Rachlin and Richard J. Herrnstein, the coeditors of this volume, gave considerable time and effort toward making the manuscript more readable. Research was supported by grants to Richard J. Herrnstein, by grants from the Dare Association, Inc., to the first author and a National Science Foundation Fellowship to the second author.

a. A simpler hypothesis would be that the subject maximizes the expected value of reinforcement, measured, say, in milligrams of grain received. However, the reinforcing value to the subject of a food pellet need not be proportional to its weight, and so a rationally maximizing subject need not regard the prospect of a 50 percent chance of receiving 60 milligrams and a 50 percent chance of receiving 20 milligrams as equal in value to a certainty of receiving 40 milligrams of grain. In order to take this into account, it is assumed instead that the subject maximizes the expected value of the utility of the reinforcement, where the utility is taken to be some monotonically increasing (but not necessarily linear) function of the amount of grain received.

events. These two phenomena can explain the observed deviations from performance values predicted by full-information aggregate maximization. Whether or not the local maximizing model that contains these two hypotheses is good, however, it is easy to support and difficult to reject with aggregate data. This is because the values of the model's parameters relating to discriminability and forgetting cannot be directly estimated from aggregate level observation. They appear as free parameters that may always be adjusted to make the theory's predictions roughly consistent with the aggregate data. The predictions generated in this fashion are not easily disproved; accordingly, confirmation of the theory is trivial.

The studies reported in this chapter seek to determine the reinforcement value of sequences of reinforcers at the local, nonaggregated level. Value is measured using two kinds of choice procedures. In the first procedure, called discrimination (Experiments 1 and 2), the subject's choice is reinforced, S^{R+}, for correctly determining from which of two reinforcement schedules, one rich and one lean in reinforcement, a given reinforcement sequence has been taken. In Experiment 1, one of sixteen different sequences of reinforcers was presented on a trial; in Experiment 2, one of 64. For each different sequence of reinforcers the proportion of trials on which the subject indicates that the richer density schedule has been sampled is separately recorded. This proportion reflects the richness or *perceived value* of that sequence. This procedure makes it possible to examine how sequences of reinforcers are valued at the local level.

In the second procedure, called preference (Experiment 3), the subject allocates its time and responses between two keys in the first link of a concurrent chain schedule. One of the second links of this schedule consists of a *comparison* sequence of reinforcers and the other a *standard* sequence of reinforcers; the sequences used here are the same 16 sequences used in Experiment 1. In the comparison link, the sequence of reinforcers is changed to a new sequence only after its value has been determined. In the standard link, the sequence of reinforcers is always the same. The number of milligrams of food reinforcement delivered within the standard link is varied until the amount required to make the subject indifferent between the two second links is found. This value that reflects *effective value* is recorded for different sequences of the comparison schedule.

Both the discrimination and the preference procedures allow sequences of reinforcers to be scaled as to relative aggregate reinforcement provided, with the discrimination experiments indicating the perceived value and the preference experiment indicating the effective value. It will be shown that these two measures of aggregate reinforcement value coincide closely.

In addition, because the value of each sequence of reinforcers is separately measured, it becomes possible to determine the disaggregated effect of each individual reinforcer according to its position in the sequence. The more sophisticated local maximizing hypothesis just described yields quite specific predictions

as to what the relative reinforcement values of different sequences should be. The assumption of a forgetting process is no longer a rationale for the introduction of ad hoc parameters, as in the case of aggregate data, since the precise nature of the forgetting process may be determined by comparing the aggregate values of different sequences of reinforcers. Consistency of the data with the sophisticated local maximizing hypothesis is no longer trivial. This hypothesis will be shown to account in detail for the relative perceived values, and, since the effective values have the same structure as do the perceived values, it is believed that a similar maximizing model can account for the effective values. Therefore, apparent matching in aggregated data may be accounted for by a local maximizing model allowing for imperfect discrimination and memory loss.

EXPERIMENT 1: DISCRIMINATION PROCEDURE

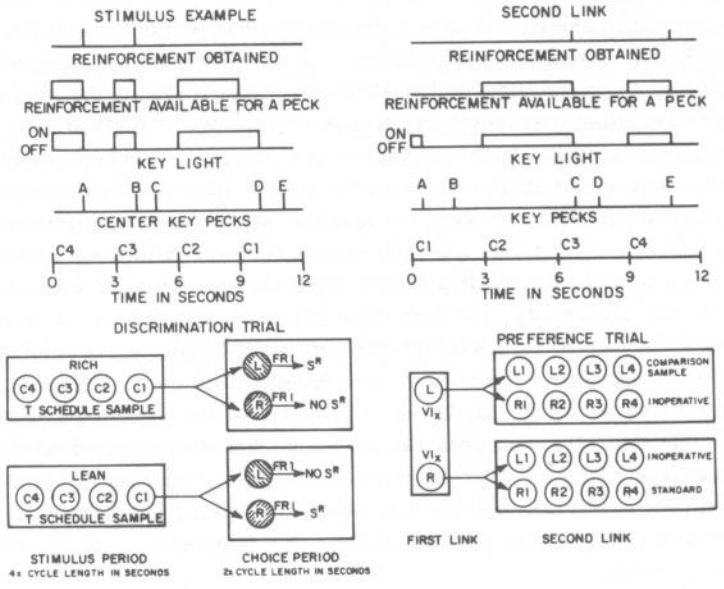
Each of four White Carneaux pigeons was run in one 256-trial session per day. Trials consisted of a stimulus period followed by a choice period, as shown at the bottom left of Figure 2-1. The task was similar to one in which the subject identifies whether a randomly chosen urn is the rich one with three reinforcer balls and one nonreinforcer ball, or whether the urn is the lean one with one reinforcer ball and three nonreinforcer balls. On each trial there are four draws from that trial's urn; these draws take place in the stimulus period. After a ball is drawn it is put back into the urn. After the four draws from the urn presented on that trial, the subject indicates whether the sample came from the rich or the lean urn; this takes place during the choice period.

The stimuli presented during the stimulus period are to be described *macroscopically* (the whole stimulus) and *microscopically* (the parts of the stimulus). First there will be a general description of the procedure, including the stimulus and choice periods, and then the contingencies in these periods will be described more specifically.

The stimuli to be discriminated in the present study were samples from one of two modified T schedules (Commons, 1979; Schoenfeld and Cole, 1972; Schoenfeld, Cumming, and Hearst, 1956; Weissman, 1961). The rich schedule had the reinforcement S^{R+} probability for a lighted center-key peck, C, of $p(S^{R+}|C) = 0.75$. The lean schedule had a $p(S^{R+}|C) = 0.25$. A sample from a schedule consisted of 4 cycles, which in the standard case were each 3 seconds in length. At the beginning of each cycle, the center key was illuminated. The cycles, c_i or just i, were numbered so that c_4, occurring at the beginning of the stimulus period, is the furthest from choice, and c_1, occurring at the end of the stimulus period, occurs just before the choice period. The first center-key peck in each of the four cycles darkened the center key and was reinforced with a probability of 0.75 (if the rich schedule was sampled) or 0.25 (if the lean schedule was sampled). Although the reinforcement probability was neither 0 nor 1

Figure 2-1. A stimulus example from a discrimination trial shown (*top left*) by means of a state diagram for a sample from a 3-second 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 it would be very unusual because there is a cycle without a peck occurring in it. The contingencies during an entire trial are shown (*bottom left*). The stimulus period contains substimuli of the form shown in the top left portion. The choice period contain a single reinforcer for correct choice or extinction for incorrect ones. The choice period immediately follows the stimulus period. There is no intertrial interval.

The second-link reinforcing stimulus from a preference trial (*top right*) is the mirror opposite of the stimulus example shown on the left. It illustrates what may happen if side-key pecks occur (and do not) occur when reinforcement has been programmed (or not) on cycles during the second link. In reality it would be very unusual because of the cycle with no peck occurring in it. The contingencies during an entire trial are shown (*bottom right*). The choice period consists of a concurrent VI 12-second, VI 12-second schedule. The first response after the interval completes the VI requirement. When the left requirement is met, the left-key changes color and has the comparison sample substimulus programmed on it. Likewise, when the right requirement is met, the second-link standard substimulus is presented on the right key, which also changes color. There is no intertrial interval.



on a cycle i, a center-key peck was ($v_i = 1$) or was not ($v_i = 0$) reinforced on each of these 3-second cycles. Only the first center-key peck in any one cycle was ever reinforced.

An example of what happens in a reinforcement schedule sample presentation, during the stimulus period of a trial, is shown in the upper left of Figure 2-1 for the substimulus 1110 (reinforcement programmed for c_4, c_3, and c_2, but not c_1). Each cycle began with the illumination of the center key. Again, this happens in all the samples. In the fourth cycle before choice (c_4) the first center-key peck, labeled A, darkened the key and was reinforced. Likewise at the beginning of the third cycle before choice (c_3) the center-key light came on and the first center-key peck (labeled B) darkened the key and was reinforced. The other center-key peck during the same cycle, labeled C, was not reinforced. At the beginning of the second cycle before choice (c_2) the center-key light came on again. There was no center-key peck during that cycle (an unlikely occurrence). Therefore the center-key light remained illuminated into the first cycle before choice (c_1), and the programmed reinforcer was not picked up. At the beginning of the first cycle before choice (c_1), the center-key light just remained on. The first peck in that cycle, labeled D, darkened the key. There was no reinforcement programmed for this trial, so none was obtained. The next center-key peck, E, had no effect since it was the second one in this cycle. For the trial as a whole, only two out of three reinforcers were obtained. Although this sample could have come from either the rich or the lean schedule, it was more probable that it came from the rich schedule because it had three reinforcements programmed.

At the onset of the choice period the side keys were illuminated; the center key stayed dark or was darkened in those rare cases where no key peck occurred in the last cycle of the substimulus. Duration of the choice period was always twice the standard or base length cycle, as shown in the lower left of Figure 2-1. The first side-key peck, whether correct or not, darkened both keys, and no further pecks were counted. If a substimulus sample from the rich schedule had been presented on the center key, the first left-key peck was reinforced (fixed ratio-1, FR-1, a hit or left correct); a right was not reinforced (extinction, EXT, a miss or right error). If a sample from the lean schedule had been 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 three different standard cycle lengths were used—2 seconds, 3 seconds, and 4 seconds ($T = 2, 3, 4$)—only the data for the 3-second standard are reported here. Each was run until the birds stabilized. In daily sessions the standard (base) cycle length was of standard length (T seconds) on 224 trials, doubled (T multiplied by 2) on 16 probe trials, and tripled (T multiplied by 3) on another 16 probe trials. The position of the probe trials within a session was randomly distributed.

Stimulus Description

Division of the stimuli into four levels of analysis is crucial for understanding the relation between stimuli at the aggregated level and stimuli at the more disaggregated levels. There are four levels of description of the *stimuli*:

1. On a *macro* level each probabilistic reinforcement schedule, either the rich ($p = 0.75$) or the lean ($p = 0.25$), is viewed as a single *stimulus*, here called S_{rich} (S_R) and S_{lean} (S_L), respectively; sampling considerations are of no concern. These two *stimuli* are equivalent to cued random-interval (RI) or variable-interval (VI) schedules with a T-second limited hold and average intervals equal to $4/3T$ second (rich) and $4T$ seconds (lean), where $T = 3$ seconds for the standard case being considered. On a variable schedule (VI) after a certain amount of time (interval) the first response is reinforced.

2. At the *molar* level all samples from the two reinforcement schedules (called at this level *substimulus samples*), with the same number of reinforcers are viewed as the same *stimulus*. Each substimulus S_n has a reinforcement density $D(S_n)$ equal to N_d, the number of center-key-peck reinforcement opportunities over the four cycles in the substimulus. Density, the defining characteristic of a stimulus at this level of analysis, ranges from 0 to 4 reinforcers per substimulus, $D(S_n) = N_d = 0, 1, 2, 3, 4$, giving rise to 5 such *stimuli*, designated S_{N_d}, as shown in Table 2-1. This level of stimulus description distinguishes among the number of reinforcements within a substimulus; it does not distinguish among substimuli with the same number of reinforcers but with those reinforcers programmed on different cycles within the $4T$-second stimulus period. The frequency of the occurrence of the possible samples with a given number of reinforcers is shown in the left panel of Figure 2-2 and in Table 2-1. There are 16 possible sequences of reinforcement or nonreinforcement in a 4-cycle sample. The probability that a particular set of 4 events will occur was obtained by expanding the binomial $(p + q)^4$ with $p = 0.75$ or 0.25 and $q = 1 - p$. Actual presentation frequencies deviated somewhat from the expected frequencies because of sampling, as shown in Table 2-1. The modal and mean density of the rich schedule is 3; the modal and mean density of the lean schedule is 1.

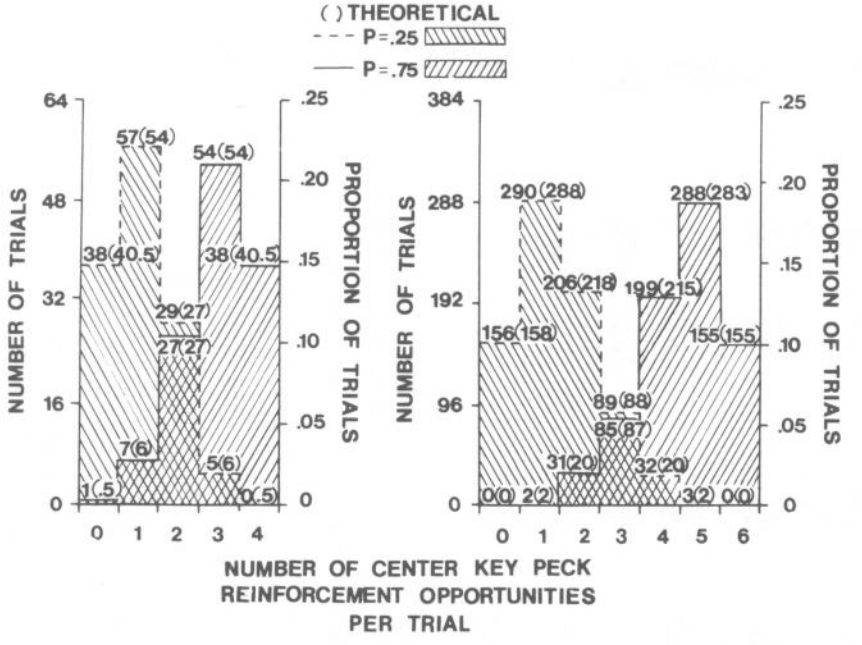
3. At the *molecular* level the definition of a *stimulus* involves the sequence or pattern of reinforcement in a sample. At this level substimuli are described not with a one-digit density number but with a four-digit binary number. A number like 0001 indicates a sample with 3 cycles without reinforcement opportunities followed by 1 cycle with a reinforcement opportunity. The substimuli S_n are numbered from 0000 to 1111. The leftmost digit represents the cycle furthest in time from choice c_4, and the rightmost represents the cycle immediately before choice c_1.

Table 2-1. Description of stimuli. For each of the five possible molar classifications of substimuli with number of reinforced cycles, N_d, 0–4, and each of the 16 possible molecular substimuli, S_n, 0000–1111, the theoretical and sampled probabilities of reinforcement are given for the rich (LC), $p = 0.75$, and the lean (RC), $p = 0.25$, schedules. The sum of the molecular frequencies equals the molar frequencies. On a molecular level 0 indicates a nonreinforced cycle and 1 a reinforced cycle, the digit on the right being closest to choice. The theoretical and sampled numbers of occurrences of molar classified substimuli and molecular substimuli in the 256 trials from a session are also given. The last column shows the probability of a left-key peck or a right-key peck being correct. This is the same as the probability of reinforcement for a left-key peck and a right-key peck.

Molar Level: Number of Reinforcers in a Substimulus N_d	*Molecular Level: Substimulus Number in Binary Notation*	*Gross Molar*						*Probability of a L or R Being Correct* $p(LC\|S_{N_d})$ $(p(RC\|S_{N_d}))$
		Rich Frequency (LC)		*Lean Frequency (RC)*		*Frequency of Corrects (LC + RC)*		
		Theoretical	*Sampled*	*Theoretical*	*Sampled*	*Theoretical*	*Sampled*	
0	0000	0.5	1	40.5	38	41	39	.0256
		0.5	1	40.5	38	41	39	(.9744)
	1000	1.5	1	13.5	17	15	18	
	0100	1.5	1	13.5	16	15	17	
	0010	1.5	3	13.5	14	15	17	
	0001	1.5	2	13.5	10	15	12	
1		6.0	7	54.0	57	60	64	.2093
								(.8907)
	1100	4.5	3	4.5	4	9	7	
	1010	4.5	5	4.5	6	9	11	
	1001	4.5	6	4.5	4	9	10	
	0110	4.5	3	4.5	5	9	8	
	0101	4.5	6	4.5	6	9	12	
	0011	4.5	4	4.5	4	9	8	
2		27.0	27	27.0	29	54	56	.4821
								(.5179)

	1110	13.5	17	1.5	2	15	19	
	1101	13.5	14	1.5	1	15	15	
	1011	13.5	11	1.5	0	15	11	
	0111	13.5	12	1.5	2	15	14	
3		54.0	54	6.0	5	60	59	.9152 (.0847)
	1111	40.5	38	0.5	0	41	38	
4		40.5	38	0.5	0	41	38	
		128	127	128	129	256	256	1.0000 (.0000)

Figure 2-2. The distribution of reinforcement for the first center-key peck ($S^{R+}|C$) for 4-cycle (*left*) and 6-cycle substimuli (*right*) 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 for the 4-cycle substimuli and 0.767 and 0.233 for the 6-cycle substimuli.



4. At the *micro* level the definition of a *stimulus* depends on whether or not there is a reinforcement opportunity on a particular cycle c_i, i cycles before choice, irrespective of what is programmed for its neighbors. There are 8 such *micro* stimuli (4 with reinforcers on each of the 4 cycles and 4 without reinforcers on each of the 4 cycles).

A molecular substimulus, even 0000 or 1111, could be sampled from either schedule stimulus. What differed therefore was not the substimulus that could be sampled from the schedule stimuli but the frequency with which a substimulus came from the rich schedule or the lean schedule. These frequencies are presented in Table 2-1. A comparison of columns 1 and 2 with the last column in the table shows that the greater the number of cycles in a substimulus having a

reinforcement opportunity, the greater the likelihood that the rich schedule was in effect. Therefore, the rich schedule was more likely to be in effect when 3 or 4 reinforcement opportunities were programmed, and the lean schedule was more likely to be in effect when 0 or 1 reinforcement opportunities were presented on a trial. Substimuli with 2 reinforcement opportunities occurred about equally often given either schedule. Again, when the rich schedule was in effect $p(S^{R+}|L)$ increased. Birds were run until their performances stabilized. The data reported here were collected in the 5 sessions following stabilization.

EXPERIMENT 2: SIX 3-SECOND CYCLE SAMPLES

Method and Procedure

The subjects and apparatus for this experiment were identical to those described in Experiment 1, except that one less subject was used. The discrimination task was similar to that used in Experiment 1, except that the number of cycles was increased from 4 to 6. The two schedules to be discriminated were, again, two overlapping distributions (stimuli) of reinforcement opportunities, one with $p(S^{R+}|C) = 0.767$ (the rich schedule) and one with $p(S^{R+}|C) = 0.233$ (the lean schedule). These probabilities will be treated as approximately the same as for the 4-cycle experiment for purposes of later analysis. The stimulus on each trial was one of the 64 substimuli that could be generated by taking samples of size 6 from one of the two T schedule stimuli. The frequency of occurrence of the possible samples with a given number of reinforcers for the 6-cycle case is shown in the right panel of Figure 2-2. The modal density of the rich schedule is 5 and the lean schedule is 1; mean value can be obtained by multiplying the probability of occurrence of a reinforcer times the number of cycles (e.g., $0.767 \times 6 = 4.60$ for the rich and 1.40 for the lean). The value of T was again equal to 3 seconds, and there were no probe trials.

EXPERIMENT 3: PREFERENCE PROCEDURE

Each of ten White Carneaux pigeons was run in one 128-trial session per day. A trial consisted of the first and second links of a concurrent chain schedule as shown in the bottom right-hand portion of Figure 2-1. In the top half of Table 2-2, a simple version of a concurrent chain schedule is shown. In a concurrent schedule two schedules are in effect at the same time, one on each key. The center key is inoperative. The bird pecks either the right or the left key. The second link on either key is activated by the completion of the contingency on the first link. Both keys are freely available during the first link, but once the bird has entered the second link on either key the other key is deactivated. In this simple version the first link on both the right and the left keys is a 12-second variable-

Table 2-2. Simple concurrent schedule compared with concurrent chain schedule used here.

Simple Concurrent Schedule

VI 12 seconds ⟶ Reinforcer

VI 12 seconds ⟶ Reinforcer

Concurrent Chain Schedule

VI 12 seconds ⟶ v_1 v_2 v_3 v_4 on left comparison key

VI 12 seconds ⟶ 0 V_{mg} 0 0 on right standard key

$$v_i = \begin{cases} 0 \text{ milligrams} \\ 20 \text{ milligrams} \end{cases}$$

V_{mg} = varies

interval schedule. The second link is a single reinforcer. Completing the VI 12-second requirement on either of the keys would lead to the obtaining of the reinforcer programmed for that key. In the more complex case shown both at the bottom right of Figure 2-1 and bottom of Table 2-2 the single reinforcer of the second link is replaced by one of two samples from a T schedule. Therefore the second link's T schedule sample on a given key serves as the reinforcer for completing the requirement on the same key during the first link. Presumably for this example different values of the reinforcers in the second link would lead to differential responding on the left and right keys. The samples in the second link are the same as those used in the 4-cycle discrimination experiment (Experiment 1), so that the aggregation of sequences of reinforcers in the two kinds of experiments may be directly compared. Again, a schematic diagram of the first and second links is shown in the bottom half of Table 2-2.

In order to understand the contingencies in the second link let us return to a short description of the reinforcement samples. Recall that the reinforcement samples in Experiment 1 consisted of four 3-second cycles. A reinforcer is either programmed or not on each cycle. A total of 16 such samples, called substimuli, are possible. The 16 substimulus samples each consisted of four 3-second cycles. Each cycle began with the illumination of the left key, and the first left-key peck was ($v_i = 1$) or was not ($v_i = 0$) reinforced. Subsequent pecks during a cycle were not reinforced. The cycles c_i were numbered so that the c_1 was the cycle at the beginning of the second link right after choice, and c_4 is the cycle furthest from choice occurring at the end of the second link. Notice that this is exactly the same as was done in Experiment 1. In both cases c_1 is the cycle closest to choice and c_4 the cycle furthest away; what changes is where the choice occurs relative to the four cycles. For the preference experiment, however, the binary-numeral notation for substimuli will be reversed. That is, 0111

means $v_1 = 0$, $v_2 = 1$, $v_3 = 1$, $v_4 = 1$. This is so that left to right will continue to indicate temporal ordering. An example of an unlikely response pattern to the 0111 substimulus sample is shown in the upper right of Figure 2-1. At the beginning of the first cycle, c_1, the key was illuminated. The first left-key peck darkened the left key and was not reinforced. Although the reinforcement was set up during the second cycle and in the actual experiment the bird would always peck and obtain it, for the sake of illustration, a cycle without a peck is shown here. Because there was no peck, no reinforcer was delivered during that cycle and the light remained on. At the end of that cycle, the reinforcer was set up again. In the third cycle after choice the bird did peck and obtained the reinforcer; the key immediately darkened. The same thing happened during the fourth cycle, c_4.

Two aspects of these reinforcement samples serve as independent variables: (1) the pattern of reinforcement, that is, on which of the four cycles reinforcers were delivered, and (2) the amount, in milligrams, of the reinforcer delivered on a particular cycle. On the left key, the *comparison* key, the pattern of reinforcement varied while the amount delivered on a cycle stayed constant. On the right key, the *standard* key, the reinforcement amount was varied but the pattern was always the same one. The conditions determining when these dimensions of the reinforcement samples were varied will be described next.

On the left, comparison key, the peck that completed the VI led to the comparison substimulus. The same comparison substimulus was present for a minimum of 30 sessions. The maximum number of sessions it was run was determined by conditions discussed later. Throughout the study all reinforcers obtained by comparison-key-pecks in a cycle were 20 milligrams.

On the right, standard key, the peck that completed the VI led to the standard substimulus, $0V_{mg}00$. In this substimulus no reinforcers were programmed on the first, third, and fourth cycles. Reinforcement, consisting of varying number of pellets, was always programmed on the second cycle. The amount of reinforcement, V_{mg}, during the second cycle was varied by the experimenter. The number of sessions that had a given value V_{mg} programmed depended on conditions to be outlined. To get values such as 30 milligrams using just 20-milligram pellets, 0.5 of the c_2 cycles would have two pellets and 0.5 would have one pellet. A similar probabilistic delivery of reinforcement was used to generate values smaller than 20 milligrams. In this second case, the hopper light still went on when a reinforcer was to be earned, but the pellet was delivered probabilistically to generate approximately any required value between 0 and 20 milligrams.

The treatment of probabilistic reinforcement as equivalent to certain delivery of the mean reinforcement amount implicitly assumes that the utility obtained from reinforcement is a linear function of amount. In fact, as will be shown, utility is a slightly concave function of the number of pellets. But since the exponent obtained for the power function is not too much less than one (approxi-

Table 2-3. Plan for the analysis, with data from the preference study compared with data from the discrimination study for both the perceived value and the effective value (there is no model in the effective value and the models in the elemental data). Column 1 shows the substimuli in preference data in digital form and Column 2 shows the same substimuli in binary form detailing the reinforcement pattern. The equivalent substimuli in the discrimination case run just the reverse (0001 in the preference case is equivalent to 1000 in the discrimination case). The cycles are read from left to right; a 0001, having the reinforcer, is the fourth cycle after choice in the preference case. The titrated values in milligrams are shown for each bird that was tested on a given comparative substimulus; brackets indicate that the bird eventually died after the particular data were being collected. The mean value in milligrams is shown (Column 4) along with its standard deviation.

Preference			*Discrimination*					
Substimulus Number								
Digital	*Cycles*	*Titrated* V_{mg}	$\bar{V}_{mg}$ *(S.D.)*	$\bar{Z}_{p(L)}$ *(S.D.)*	$Z_{p(L)_1}$	$Z_{p(L)_2}$	$Z_{p(L)_3}$	$Z_{p(L)_4}$
00	0000		0	−2.420 (.30)	−2.035	−2.750	−2.371	−2.517
01	0001	12_{30} $[8_{85}]$ 4_{346} 7_{58} 9_{58}	8.000 (2.92)	−2.026 (.53)	−1.492	−2.086	−2.739	−1.787
02	0010	10_{346} 14_1 10_1 12_{30}	11.500 (1.91)	−1.741 (.78)	−1.025	−2.043	−2.704	−1.193
03	0011	29_{346} 30_{346} $[30_{85}]$	29.667 (.577)	−0.874 (.84)	−0.262	−0.529	−2.108	−0.595
04	0100	20_{58} 18_{59} 20_{59} 24_{62} 18_{161} 20_{161} $[20_{85}]$ 20_{62}	20.000 (1.85)	−1.255 (.45)	−1.025	−1.155	−1.918	−0.923
05	0101			−0.596 (.61)	−0.449	−0.449	−1.460	−0.026
06	0110	40_1	40.000 —	−0.326 (.85)	0.131	−0.169	−1.565	0.299
07	0111	94_{62}	94.000	0.658 (.84)	1.396	0.804	−0.553	0.983

08	1000	44_{59}	44.000 —	−0.783 (.44)	−0.470	−0.430	−1.383	−0.849
09	1001	$[80_{85}]$	80.000 —	0.05 (.77)	0.288	0.537	−1.097	0.472
10	1010			0.290 (.64)	0.272	0.734	−0.609	0.764
11	1011	$[70_{85}]$ $[50_{84}]$	60.000 (14.14)	1.216 (.79)	1.684	1.534	0.038	1.609
12	1100	$[70_{84}]$	70.000 —	0.900 (.65)	0.863	1.698	0.111	0.929
13	1101			1.794 (.84)	1.696	2.536	0.657	2.287
14	1110	$[70_{84}]$ $[82_{84}]$	76.000 (8.48)	1.820 (.50)	1.732	1.960	1.192	2.394
15	1111	144_{30} 153_{346} 160_{161} $[80_{102}]$	134.250 (36.75)	2.386 (.51)	2.983	2.276	1.754	2.532

mately 0.7), this method of interpolating values is reasonably accurate over the range of values in question. An extrapolation of Mazur's (1981) results shows that probabilistic delivery of reinforcement is a reasonable procedure for obtaining intermediate amounts.

Mazur, using a Stubbs–Pliskoff procedure (a concurrent interdependent procedure), showed that probabilistically delivered reinforcement, $p(S^{R+}) = 0.50$ was valued as 0.50 and $p(S^{R+}) = 0.33$ was valued as 0.33 (in his procedure relative value was measured by relative rate). In Mazur's study, relative left-key pecking rate, Rel B_L (his measure of relative value) is equal to the rate of left-key pecking, B_L, divided by the sum of the left-peck rate and the right-peck rate, B_R, Rel $B_L = B_L/(B_L + B_R)$. This same relation, between probabilistically delivered reinforcement and relative rate (as defined for this situation) should hold in the present situation, which uses a concurrent chain schedule (instead of a simple concurrent schedule) where one of the pellets delivered in the second link's standard substimulus was probabilistically delivered as described before. The probabilistic reinforcement p can be used to obtain a reinforcement amount reasonably equivalent to the actual milligram value desired, as follows: For m = number of pellets,

$$V_{mg} = m(20 \text{ mg}) + p(\text{reinforcement}) \times 20 \text{ mg} = 20(p + m) \text{ mg}$$

For example, to obtain a reinforcement amount equal to 10 mg, m would equal 0 and p would equal 0.5, and therefore $V_{mg} = 0.5 \times 20 \text{ mg} = 10 \text{ mg}$.

Three kinds of information were recorded: first, rates of left-key and right-key pecking during the initial links; second, which comparison substimulus was programmed on the left key during the second link, and third, the amount of food, V_{mg}, delivered for the reinforced right-key peck in the second link standard substimulus $0V_{mg}00$. The amount of food delivered in the second link standard substimulus was varied in order to get the titrated amount of food. The titrated amount of food was defined as the number of milligrams of food necessary to produce indifference between the comparison and the standard. Such indifference was defined as a relative rate of left-key pecking equal to 0.5 ± 0.1 for 6 consecutive sessions. There also had to be 6 consecutive sessions for food amounts that produced relative rates above the 0.5 indifference value and 6 consecutive sessions below.

The titrated amount of food V_{mg} was not the only parameter that was systematically varied. After V_{mg} had been found for a particular substimulus, the substimulus that served as the comparison substimulus was changed and a new V_{mg} was obtained for the new substimulus. A given substimulus for which V_{mg} is found is referred to as a point. This procedure was not carried out using all substimuli with all birds. Which birds were run on which substimuli is shown on the left side of Table 2–3. Next to each substimulus number (listed in the column entitled cycles) is a list of the titrated amounts obtained for that substimulus by each bird. Entries are in the following form: Titrated amount$_{\text{bird number}}$. Thus,

Table 2-3a. Organization of the presentation of the data, analysis, and corresponding theory.

	Elemental Data	*Transformation*		*Additivity*		*Decrementing*	
Perceived Value		Exp.	Formal Theory	Exp.	Formal Theory	Exp.	Formal Theory
		X	X	X	X	X	X
Effective Value		X	X	X	X	X	

the entry 12_{30} under substimulus 0001 means that bird 30 is indifferent between V_{mg} = 12 mg and substimulus 0001. Numbers in brackets are for birds that died in the course of the study; three of the ten birds died during the study, which took over three years to run. Some of the points were replicated by the same bird.

Results of the Three Experiments

The analysis is presented in four parts: (1) elemental data for perceived value and effective value of a substimulus; (2) transformations that make the value of the molar level *stimuli* a linear function of the reinforcement density; (3) the degree to which the aggregation of reinforcers within a substimulus is a linear function of the contribution provided by each reinforcer when the transformations from (2) are applied; and (4) the decrementing functions that best account for the decrease in value as a reinforcer occurs further from choice. (See Table 2-3a.) The analysis will be presented first with the corresponding model following.

ELEMENTAL DATA

The Discrimination Situation

The most elemental form of the discrimination data examines the choices that the pigeons make following a substimulus presentation. A left peck to a substimulus, $L|S_n$, is the decision that indicates that the substimulus is rich. The probability of a left-key peck to the given substimulus is $p(L|S_n) = (L|S_n)/(L|S_n + R|S_n)$. This value is called bias in signal detection theory (Nevin, 1969; Green and Swets, 1966), the model used here and in previous presentations of similar data (see Commons, 1979; 1981). This represents the tendency to indicate that the substimulus comes from the rich schedule. Although only single left-key peck or right-key peck dichotomous choice can occur on each trial,

$p(L|S_n)$ indicates the proportion of times the pigeon indicates that the sample was from the rich schedule, or in the following sense, the perceived richness of the sample.

It is assumed that probabilistic responding is a result of the combination of signal and random noise in the perceptual and memory processes for reinforcers in the schedule sample. For a substimulus whose perceived density is very high when measured over a number of trials, the perceived density on a trial will be high no matter whether the noise processes increase it or decrease it, unless the noise is extremely large, and so the subject will almost always peck left (high $p(L|S_n)$). Similarly, for a substimulus whose perceived density is very low when measured over a number of trials, the perceived density on a trial will be low no matter what the sign of the noise terms, and so the subject will almost always peck right (low $p(L|S_n)$). When the substimulus has an intermediate mean perceived density, it will sometimes be perceived as high density and sometimes low density, depending on the noise terms, and so $p(L|S_n)$ will have an intermediate value. Furthermore, $p(L|S_n)$ will be higher the higher the mean perceived density, and hence $p(L|S_n)$ can be used as an indicator of the relative perceived density of different substimuli.

Decision Rules

The quantity $p(L)$ versus some stimulus dimension, $p(L|\text{“}S\text{”})$ is a decision rule. A decision rule states that the conditional probability of a particular choice given that certain situations or parameters of situations have occurred. In the discrimination studies it describes the control of choice by various parameters of a substimulus, such as reinforcement amount and proximity of reinforcement opportunity to choice.

Decision rules express a functional relation similar to the one between relative response rate and relative reinforcement rate in concurrent schedule research. They will be used to provide a framework in terms of which various versions of maximization laws and matching laws may be examined. Whether maximizing or matching best accounts for the decision rules shown here has implications for which is a better account in general. The results here should apply to other cases because the situation is just a parametric variation of concurrent schedules.

Nevin (1969) has pointed out that experiments similar to this one can be characterized as multiple schedules. In this case the multiple schedule has two components that are concurrent schedules of the following form: conc FR-1, EXT and conc EXT, FR-1. The stimuli marking the presence of a particular multiple component, rather than being a red or green illuminated key, are different density reinforcement schedules. Since the schedules are not perfectly discriminable, the multiple schedule is to some extent a mixed schedule. This situation is similar to illuminating two yellow lights, one signaling that a left-red

is reinforced on FR-1, with the right-green key on extinction, and the other signaling that the red and green keys have the schedules reversed on them. In either case, the relative payoff probability (rate) for making a choice in the presence of a given substimulus is known and presumably the birds' choices are controlled by this probability in a similar fashion to the way choice is controlled in the more usual concurrent schedule studies. However, relative obtained reinforcement rate for each choice is *not* the same as the probability of a choice being correct.

Matching of obtained outcome is the same as maximizing of preprogrammed or obtained reinforcement in a discrete choice situation. The more the animal goes to the more highly paid-off side, the more reinforcement is picked up on that side relative to the other. A maximizing account at the molar level would predict that a decision rule that maximizes reinforcement for choice would be the following: The optimizing ideal observer always pecks right for density 0 and 1 substimuli and always pecks left for density 3 and 4 substimuli. For density 2 substimuli, a maximizing account would make no prediction since reinforcement for making either choice was essentially equal. For the 6-cycle case, the optimizing ideal observer would always peck right for densities 0, 1, and 2, always peck left for densities 4, 5, 6, and follow any rule for density 3.

The Empirical Decision Rules at the Molar and Molecular Level: A Matching of Programmed Reinforcement Account Works at the Molar and Fails at the Molecular Level. Recall that substimuli may be described on the molar and molecular levels. On the molar level only reinforcement density is considered and substimuli take on reinforcement density values from 0 to 4 (or from 0 to 6). On a molecular level, there are 16 substimuli, each indicating a particular reinforcement pattern that may occur over the 4 cycles, or 64 substimuli, each indicating a particular reinforcement pattern that may occur over the 6 cycles. The substimuli are described in binary notation. For example, substimuli from the 4-cycle case with a density value of 1 at the molar level can have molecular values of 0001, 0010, 0100, and 1000.

Figure 2-3 shows both the molecular and molar level data. The panels in the left column of the figure show data for the 4-cycle discrimination experiment (Experiment 1). The bottom three panels of the right column show data for the 6-cycle discrimination experiment (Experiment 2). The top right panel shows 4-cycle preference data. The discrimination data will be discussed first.

At the molecular level each panel (whether for the 4-cycle or the 6-cycle data) shows one bird's $p(L)$ for each substimulus density during the test sessions on that condition. Each point indicates the probability of a left-key-peck given a specific substimulus, $p(L|S_n)$.

At the molar level the decision rules are indicated by the average left-peck probability for substimuli at a given density $p(L|S_{N_d})$. These points are the

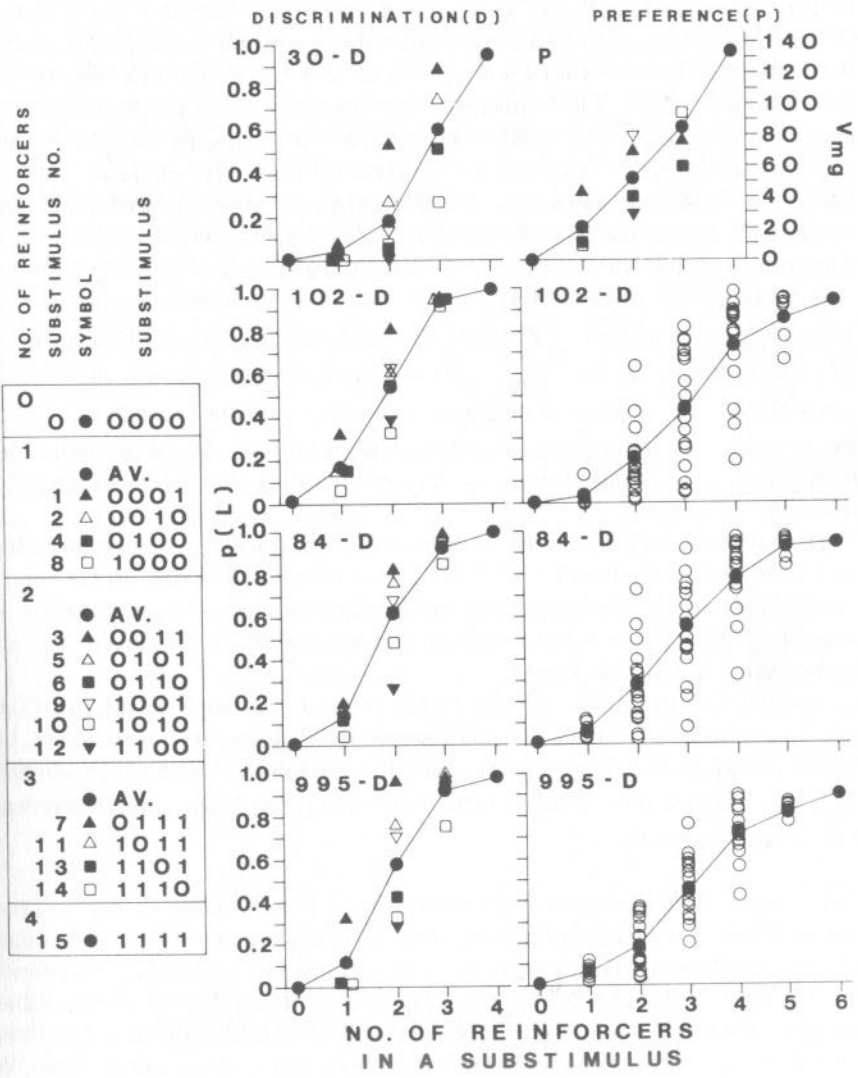
Figure 2-3. Left-key peck probability $p(L)$ for the 4-cycle discrimination data for each individual substimulus, arranged so that all same-reinforcement density substimuli are grouped above that reinforcement density number, N_d. The solid line shows the average $p(L)$ for each substimulus group. At the top of the right column is a graph of the values of V_{mg}, the amount of food in milligrams necessary to produce indifference between the standard and comparison substimuli, plotted against number of reinforcers in a substimulus. This plot, which is from the preference data, is similar to that seen for the discrimination data. The bottom three panels in the right column show $p(L)$ values for each individual substimulus for the 6-cycle discrimination data. Since there are 64 such substimuli, they are not identified individually here. The average $p(L)$ for each substimulus group is shown by the solid line.

filled-in circles. The curve through these points is close to one where the probability of left is equal to the probability of programmed payoff for that choice. The shape of the curves appears to be ogival.

The panels of this figure show clearly that choice probabilities at the molecular level, reflecting the value of substimuli, do not generally fall close to the programmed reinforcement matching curve, $p(L|S_{N_d})$. This suggests that same reinforcement density substimuli are not given equivalent value on a molecular level, although they are equivalent on a molar level by definition. That is, at the molecular level, substimuli with the same number of reinforcers in them act differently in determining the left-key-peck probability within the discrimination procedure.

A Simple Maximizing Account and Matching of Obtained Reinforcement Do Not Account for Empirical Decision Rules at the Molar and Molecular Level. Neither molar nor molecular maximizing of payoff or matching of obtained payoff are supported by the discrimination data in Figure 2-3. Remember that the assigned values for substimuli should be the minimum for substimuli with densities below the mean (0 and 1 in the 4-cycle case and 0, 1, and 2 in the 6-cycle case and should be maximum for values above the mean (3 and 4 in the 4-cycle case and 4, 5, and 6 in the 6-cycle case). In none of the panels of Figure 2-4 is this maximization nor the equivalent obtained payoff matching indicated. Neither the average nor the individual values for substimuli (for densities other than 2) fall on the discontinuous maximizing function, $p(L|S_{N_d} = 0 \text{ or } 1) = 0$, $p(L|S_{N_d}) = 2$ = any value, and $p(L|S_{N_d} = 3 \text{ or } 4) = 1$. With respect to matching to obtained payoff, there is strong undermatching. This would be consistent with the data of Chung and Herrnstein (1967) since reinforcers do not immediately follow the end of the first link. They found that different delays produce different values. A matching of obtained reinforcement account is consistent with the data but does not account for it.

Figure 2-3


DISCRIMINATION(D)
PREFERENCE(P)
30 - D
P
102 - D
102 - D
84 - D
84 - D
995 - D
995 - D
p(L)
Vmg
NO. OF REINFORCERS IN A SUBSTIMULUS
NO. OF REINFORCERS
SUBSTIMULUS NO.
SYMBOL
SUBSTIMULUS
0 ● 0000
1 ● AV.
1 ▲ 0001
2 △ 0010
4 ■ 0100
8 □ 1000
2 ● AV.
3 ▲ 0011
5 △ 0101
6 ■ 0110
9 ▽ 1001
10 □ 1010
12 ▼ 1100
3 ● AV.
7 ▲ 0111
11 △ 1011
13 ■ 1101
14 □ 1110
4
15 ● 1111

ELEMENTAL DATA

The Preference Situation Obtaining the Titrated Value

In the preference case the value V_{mg} of a comparison sample is equal to the amount of food needed to titrate the value of the substimulus $0V_{mg}00$ in which the reinforcers are delivered the second cycle into the second link. Remember that the value of $0V_{mg}00$ is titrated when the relative rate Rel B_L is approximately 0.5, showing near indifference between the standard $0V_{mg}00$ sample and the comparison. However, the amount of food that produced exactly Rel $B_L = 0.5$ was not possible to find through an adjustment procedure. To obtain a more precise value in milligrams V_{mg} that would produce indifference, a regression equation was used to interpolate. Along the way to finding the titration amount, relative rates Rel B_{L_k} were found for different amounts in the standard, V_{mg_k}, on a session k. The independent variable in the regression equation was reinforcement amount V_{mg_k}, and the dependent variable was the relative rate Rel B_{L_k}. The number of milligrams necessary to produce exactly a 0.5 relative rate for both the comparison substimulus and the standard substimulus was then found. These values were very close to the values that produced near 0.5 relative rates.

Data analyzed in the preference situation included each V_{mg} at each point for each bird, as shown on the left of Table 2-3. There were 33 cases. All the points except 0000, 0101, 1010, and 1101 were run at least once. The point 0000 produces a value of $V_{mg} = 0$ but cannot be run because all the reinforcer amounts in the second link would be zero.

Another way to think about the data is to keep in mind that bias, or $p(L)$, and V_{mg} both reflect value. One can thereby see that a somewhat similar set of results should be found for $V_{mg}|S_{N_d}$ with the preference procedure substimulus S_{N_d}, that have the same number of reinforcers as for $p(L|S_{N_d})$, which is reflective of perceived value.

One Aspect of a Generalized Programmed Reinforcement Matching Account for the Preference Data Fails at Both the Molar and the Molecular Level. The value V_{mg} did not linearly increase with N_d (density) but was more of a power function, as shown in the top right panel of Figure 2-4. Recall that the average value for same density substimuli (molar measure) was ogival within the discrimination procedure. A suitable transformation would make each of these linear. Again, in the preference case the V_{mg} value of each comparison substimulus with the same number of reinforcers but different pattern $V_{mg}|S_N$ (molecular level) deviated from the mean effective value for that group $V_{mg}|S_{N_d}$ (molar). The order of deviations in value looks similar to that found in the discrimination procedure.

In order to evaluate matching in the preference case, we must use what is called the generalized matching account for obtained reinforcement (Baum and Rachlin, 1969). There is no equivalent way to test matching of programmed reinforcement in the preference experiment to that discussed for the discrimination case since relative rate was always set at 0.5. The preference procedure results do not support one aspect of the generalized matching of obtained reinforcement law of Baum and Rachlin (1969) if it is extrapolated to programmed reinforcement. Aggregate molar theories that average the effects of reinforcement across patterns of reinforcement and only worry about the total values V^t have suggested that total value is equal to the amount of reinforcement delivered on a given instance a, times frequency of reinforcement r (Baum and Rachlin, 1969), so that $V^t = a \cdot r$. However, the upward curvature shows that increasing amounts of reinforcement in the standard are necessary to balance increases of frequency of reinforcement in the comparison. As shown in the top right of Figure 2-3, it took more than twice as much reinforcement to balance twice as frequent reinforcement as shown by comparing the V_{mg} for density 4 and for density 2 substimuli. The molar values are indicated by closed circles and conform to a power function, which is inconsistent with the linear system required by the generalized matching law account.

At the molecular level, a generalized matching law account fails because in addition to what was just stated, the position of the reinforcer, with respect to the decision that evaluates value, matters. This, of course, would be expected (see Rachlin and Green, 1972).

Generalized Maximizing Fails to Account for Preference Data at the Molar and Molecular Level. Since value again is treated as the product of amount and frequency (Rachlin, Battalio, Kagel, and Green, 1981), a maximizing account will fail on the same grounds as just previously stated for the generalized matching of obtained reinforcement account. Neither of these findings is terribly surprising because, as Rachlin himself points out, increasing reinforcement might have marginal decreasing values.

ANALYSIS OF RESULTS

Plan of the Analysis

Since previous accounts do not account for either the molar or the molecular data by definitively showing either matching or maximizing, a different approach to explaining the data is taken here. Also accounts of undermatching of response probability to obtained payoff used unmeasured parameters. The samples are clearly being evaluated in a way different than suggested by these models. Next, models that might better approximate this evaluation process by using molecular level information in the substimuli will be examined.

For the 4-cycle discrimination procedure, Commons (1981) showed that there was a transform that would yield a measure of perceived value that equaled the sum of the contributions of each reinforcer occurring in the substimulus. This means that in a reinforcement schedule sample the value of the sample is the sum of the contributions of each reinforcer. However, a number of transforms were not tried in this previous work. The contribution of each reinforcer decreases the further away it occurs from the choice used to determine value. Only one decrementing function was tried; this left a number of questions to be answered about the way reinforcers are aggregated.

There are a number of steps in describing the processes by which reinforcing events in reinforcement schedule samples are aggregated. A description of part of the processes of how a series of reinforcers are valued, when they are to be perceived as stimuli, and when they are to be preferred as reinforcers, has four parts. The first step is to see what transformation makes additive the effect of each reinforcer on perceived value. Second, data on how the impact of reinforcers decreased as a function of how many cycles of possible reinforcement intervened between it and a choice that was used to measure value, are presented. Models for weighting the contribution of each reinforcer in a schedule substimulus is found. Fourth, for the preference data, the decrementing function that accounts for the decrease in the contribution of each reinforcer the further it occurs from choice is found. In general, all the models suggest that the value of samples are approximately additive after being appropriately transformed and weighted. The differences between the models, for both the discrimination and the preference case, are the transformation and weighting schemes proposed. After these steps have been completed for both procedures, values of substimuli obtained with the discrimination procedure are directly compared to values obtained with the preference procedure.

A Single Model Predicting Both Molar and Molecular Values

Since the average of same density substimuli was not linear in either the discrimination or the preference data, transformations that would make them linear would give information about the aggregate or molar functions. However, if the averages that appear ogival in the discrimination case and as a power function in the preference case can be explained by predicting each disaggregated, individual substimulus value, a much stronger explanation of how decisions are made will be at hand. As was shown in Commons (1981) additivity of the contributions of each reinforcer at the molecular level ensures that the most linear form of the molar relations will be obtained; that is, the relation between the average of values of substimuli with the same number of reinforcers and that number is linear. This also explains matching of programmed reinforcement probability at the molar level.

Transformations That Allow Additivity

Additivity is best shown by seeing how variance can be accounted for by a multivariate linear regression with the contribution of each reinforcer being represented by the coefficient for that position. Let the predicted value V^* equal the sum of each coefficient a_i times v_i, which indicates whether a reinforcer is present or not. A star after a V means that this represents a predicted value. The predicted value of a substimulus will be called $V^*_{S_n}$, whether this value is derived from the preference case or the discrimination case. The S_n refers to the perceived or obtained value of a substimulus number. The transformed value allowing additivity will be referred to as *perceived value* (discrimination case) and as *effective value* (preference case). If V_{S_n} is scaled simply as $p(L)$, it is highly nonlinear. In the preference case, if $V_{\text{mg}}|S_n$ is simply scaled as V_{mg}, it is also nonlinear. The first thing is to see how transformation of V_{S_n} increases the amount accounted for by a linear multiple regression. Again, this can be done for both the perceived value and the effective value.

Discrimination Experiments

Transformation Needed to Get Perceived Value. Recall that bias $p(L)$ is expected to be a monotonic function of perceived value. We find that as the number of reinforcers in the substimulus increases, $p(L)$ does not linearly increase. A good measure of perceived value for molar stimuli should linearly increase as the density (number of reinforcers in a substimulus) increases. Perceïved value V_{S_n} should be the sum of the contribution of each reinforcer, as was previously discussed:

$$V_{S_n} = T(p(L)) = a_0 + \sum_{i=1}^{n} a_i v_i \quad .$$

The letter T stands for the transformation that makes the multiple regression the most linear. The a_i are the weights applied to v_i, where $v_i = 1$ is when a reinforcer did and $v_i = 0$ is when a reinforcer did not occur on the ith cycle away from choice. From the shape of the curve relating $p(L)$ to density N_d in Figure 2-4 the transformations that might work are the probit (inverse probit transformation), the logit, and the arcsine, which is written $\sin^{-1}$. The probit transformation maps cumulative probabilities, the area under the normal curve $p(L)$ from $-\infty$ to z, into their respective z scores. This transformation is just the opposite of one in which the z score is converted into a cumulative probability, as in a z test. The logit transformation is

$$\text{logit}\,(p(L)) = \log \frac{p(L)}{1-p(L)} = \log \frac{p(L)}{p(R)} \quad ,$$

and the $\sin^{-1}(p(L))$ is the value whose sine is $p(L)$. All these transformations yield values of $p(L|S_n)$ ranging from approximately -2.5 to 2.5. A scale translation constant a_0, which is estimated in the multiple regression, is also included in the transformation, to get the values to go down to a negative value. This is done because the sum $a_i v_i$ is always positive with a reinforcement instance arbitrarily assigned a value of 1 and an instance of nonreinforcement arbitrarily assigned a value of 0.

These three transformations of $p(L)$ were tested to see which produced the best-fit regression line to the points produced by the discrimination procedure shown in Figure 2-3. The results are summarized in Table 2-4. The probit transformation did the best, as indicated by an $r^2 = 0.8175$, with the logit very close behind with an $r^2 = 0.8097$. The $\sin^{-1}$ transformation did significantly worse, with $r^2 = 0.2707$.

The sum of $a_i v_i$ was assumed to have an exponent equal to 1 after $p(L)$ was transformed. This, along with the specific function describing the decrease in weights a_i, the further a reinforcer occurs from choice, are discussed in the next section.

The Models for the Decremental Contributions of Reinforcers to Perceived Value as They Recede from Choice. Four models for weighting the values of reinforcers in the sample substimuli were tested. The four models tested for the 4- and the 6-cycle discrimination were: the multivariate linear, the exponential, the hyperbolic, and the linear. Each of the latter models may be considered a special case of the multivariate linear:

$$T(p(L)) = (a_0 + \sum_{i=1}^{4} a_i v_i)^n \quad .$$

The sum was raised to some power n to see how close it would come out to 1 and to see if there was a multiplicative effect between terms. The value of a schedule sample V is equal to the sum of the effects of a reinforcer on each cycle, all raised to some power.

Table 2-4. Three transformations of $p(L)$ fit to the multivariant model of the contribution of each reinforcer to the aggregate perceived value.

Cycles	*Discrimination Form*	*Transformation*	r^2	*df*	*MSRes*	*F*
6	Multivariate $\Sigma a_i v_i$	Probit, $z^{-1}(p(L))$	.8175	186	.213	138.9
		Logit, $\log \frac{p(L)}{1-p(L)}$	.8097	186	.234	131.9
		$\text{Sin}^{-1}(p(L))$	.2707	186	.336	11.5

The multivariate model estimates each weight a_i as a separate parameter. The other models each involve only two free parameters (in addition to a_0 and n). Muth (1960) in forecasting theory and Wickelgren (1974) have argued for exponentially weighted models, while Mazur (1981) has suggested a hyperbolic weighting. No one supports the fourth kind of model, the linear model, but since it is another simple two parameter family, it was included for purposes of comparison. The hyperbolic hypothesis has the most support in the operant literature.

As Mazur points out in the next chapter of this volume, a shift from preferring one immediate small reinforcer over a slightly larger but slightly delayed one, to preferring the reverse after an initial delay is added to both, could not occur with exponentially decreasing value. This shift means the curve for the smaller reinforcer is above the larger for short common initial delays. With longer common initial delays, the curves then cross. With an exponential decay function, the heights of the curves always have the same ratio one to another and could not cross. A hyperbolic curve, however, would cross.

In Table 2-5, the models for 4-cycle and 6-cycle cases are compared. The exponent for all the models in the 4-cycle samples is very close to 1.08 with a standard deviation of 0.14. The exponent for the 6-cycle models is 0.85 with a standard deviation of 0.82. These are not significantly different from 1.0, satisfying the use of that exponent elsewhere for discrimination data. The residual squared errors for the exponential and hyperbolic models are not very different from that found for the multivariate, showing that they well represent the weighting functions, the exponential doing slightly better in the 4-cycle case and the hyperbolic doing much better in the 6-cycle case. The linear models do not do as well; however, the differences among all the models are not significant in the range that had data. The last three models require four parameters to be estimated.

Figure 2-4 shows the degree to which the various models fit the points predicted by the multivariate regression. This simply is a way of illustrating graphically the data statistically examined in Table 2-5. Two ways that the models could be visually compared within an experiment as well as across experiments are either to compare the weights generated by the various models or to compare how well they predict substimuli that have only a single reinforcer, as that reinforcer moves further and further from choice. The latter method was used because it included the effects of the power to which the sums were raised as well as a_0, the minimum value in the discrimination cases. Instead of using actual data points, which would not reflect the entire data set, from which it would be hard to visualize the weighting functions, values for substimuli generated by the multivariate fit were used. Since each reinforcer occurred only once in any of these substimuli, the degree to which a model predicts the value of a multivariate substimulus reflects how well it estimates each independent weight for the contribution of a reinforcer on a cycle.

Table 2-5. Models for the discrimination study and the preference study. The second column gives the model's name and the third column shows the right hand side of the equations. The coefficients for these models are listed in order (Columns a_0 - a_6) with standard deviations beneath. Column *n* lists the exponent calculated for each model and standard deviation. Column *RS* lists the residual sum of squares. The rightmost Columns *df*, *MS*, and number

	Discrimination				
Cycles	*Form*	*Equation*	a_0	a_1	a_2
4	Multivariate	$a_0 + (\Sigma a_i v_i)^n$	-2.542	1.776	1.279
	S.D.		.27	.41	.30
	Exponential	$a_0 + (\Sigma a_1 v_i e^{-a_2 c_i})^n$	-2.544	2.502	.3397
	S.D.		.27	.62	.058
	Hyperbolic	$a_0 + (\Sigma a_1 v_i/(a_2 + c_i))^n$	-2.550	3.547	.9743
	S.D.		.27	1.00	.47
	Linear	$a_0 + (\Sigma v_i(a_1 + a_2 c_i))^n$	-2.548	2.090	-.3751
	S.D.		.27	.48	.10
6	Multivariate	$a_0 + (\Sigma a_i v_i)^n$	-2.599	1.696	1.339
	S.D.		.20	.36	.29
	Exponential	$a_0 + (\Sigma a_1 v_i e^{-a_2 c_i})^n$	-2.598	2.138	.2423
	S.D.		.20	.46	.024
	Hyperbolic	$a_0 + (\Sigma a_1 v_i/(c_i + a_2))^n$	-2.536	3.122	.3783
	S.D.		.19	.31	.22
	Linear	$a_0 + (\Sigma v_i(a_1 + a_2 c_i))^n$	-2.599	1.757	-.2192
	S.D.		.21	.37	.050
	Preference				
Cycles	*Form*	*Equation*	a_0	a_1	a_2
	Combined Preference				
4	Multivariate	$(\Sigma a_i v_i)^n$	—	7.290	7.394
	S.D.		—	2.60	2.08
	Exponential	$(\Sigma a_1 v_i e^{-a_2 c_i})^n$	—	11.505	.2087
	S.D.		—	5.00	.096
	Hyperbolic	$(\Sigma a_1 v_i/(c_i + a_2))^n$	—	33.300	2.407
	S.D.		—	16.25	2.04
	Refeed	$(a_1 \Sigma a_i v_i)^n$	—	7.173	—
	S.D.		—	2.44	—
	Linear	$(\Sigma v_i(a_1 + a_2 c_i))^n$	—	9.992	-1.281
	S.D.		—	3.80	.79

Table 2-5. continued
of parameters estimated. These values are the results of nonlinear regression where the independent variable predicted is the $p(L)$ of a substimulus in the preference case and the effective Value V_{mg} of the discrimination variables. The other independent variable is whether or not a reinforcer occurred on a given cycle.

a_3	a_4	a_5	a_6	n	RS	df	MS	*Number of Parameters Estimated*
.8518	.6763	—	—	1.081	22.312	58	.385	6
.22	.20	—	—	.14				
—	—	—	—	1.079	22.392	60	.373	4
—	—	—	—	.14				
—	—	—	—	1.081	22.541	60	.376	4
—	—	—	—	.14				
—	—	—	—	1.077	22.893	60	.382	4
—	—	—	—	.14				
.9000	.7261	.7393	.5442	.8490	36.387	184	.198	8
..19	.17	.17	.14	.082				
—	—	—	—	.8448	37.371	188	.199	4
—	—	—	—	.081				
—	—	—	—	.8578	37.061	188	.197	4
—	—	—	—	.081				
—	—	—	—	.8493	39.110	188	.208	4
—	—	—	—	.083				

a_3	a_4	a_5	a_6	n	RS	df	MS	*Number of Parameters Estimated*
2.395	5.945	—	—	1.561	6976.8	28	249.2	5
1.41	1.80	—	—	.15				
—	—	—	—	1.453	8324.6	30	277.5	3
—	—	—	—	.15				
—	—	—	—	1.443	8295.7	30	276.5	3
—	—	—	—	.15				
—	—	—	—	1.382	8510.2	31	274.5	2
—	—	—	—	.14				
—	—	—	—	1.468	8401.1	30	280.0	3
—	—	—	—	.15				

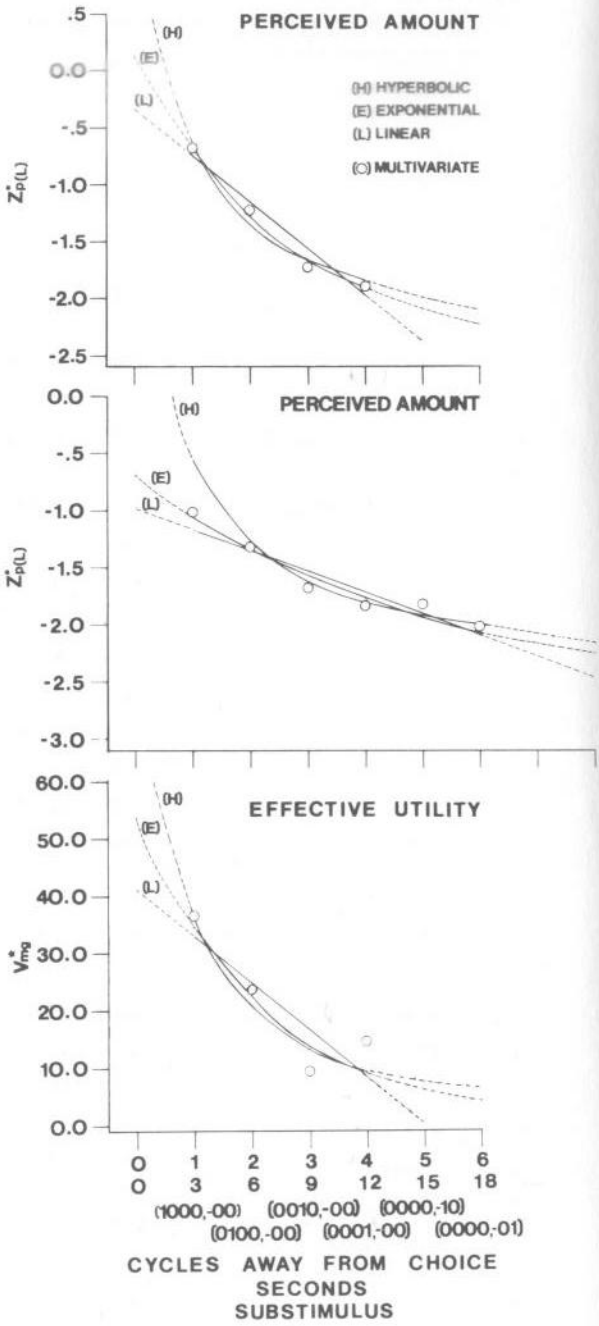


Figure 2-4. The extent to which the curves generated from the forgetting models fit values of substimuli derived from the multivariate regression estimates is shown. The curves represent the three forgetting models, hyperbolic, exponential, and linear. The points represent the predicted substimulus values. Substimuli are arranged so the one reinforcer in them recedes further from choice as one goes further to the right. The top two panels show perceived values for substimuli found from 4- and 6-cycle discrimination studies; the bottom panel shows effective values for substimuli found from the 4-cycle preference study.

A number of things are clear from an examination of this figure. First, the curves are always in the same order outside the region where the fitted points lie. The hyperbolic is the most curved and falls above the other curves, the exponential is next, and finally the linear lies below. Within the range of N_d values of the set of substimuli being predicted, the order is reversed. Second, the hyperbolic curve misses the first point in the 6-cycle case but was closer to the other points than the other curves. This suggests that with more points the hyperbolic may be even more favored. Third, the fits would seem to be better differentiated if there were more cycles. Mandell (1981) has pigeons discriminate longer schedule samples successfully, but because she used no samples of shorter length, her data could not be analyzed as was done here. The curves suggest that the different models diverge most clearly in their predictions for the value of reinforcers very close to choice. But in order to program reinforcers closer to choice, cycle length would have to be shortened. There are problems in doing that. When the cycle length was shortened to 2 seconds (Commons, 1981), there was a loss of discriminability because the cycles were occurring so close together that the birds missed picking up reinforcers.

Preference Experiment (Effective Value)

Transformation Needed to Get Effective Value. Recall that V_{mg} is expected to be a monotonic function of effective value. We find that as the number of reinforcers in the substimulus increases, V_{mg} did not linearly increase (as was shown in the top right panel of Figure 2-3). Effective value $T(V_{mg})$ should be a transformation of V_{mg} that is equal to the sum of the contribution of each reinforcer, if we follow a process that is parallel to that used for perceived value. Recall that additivity is best shown by a multivariate linear regression where the predicted variable is the transformation of V_{mg}:

$$T(V_{mg}) = V_{mg}^{1/n} = \sum_{i=1}^{4} a_i v_i \quad .$$

Only a power function was tried here. The value for $1/n$ found in the multivariate regression was 0.7. Its inverse was 1.41 with a standard deviation of 0.14, which was significantly different from 1 ($p < 0.01$). This result is reported in the bottom of Table 2-5. That exponent $1/n$ explains the curvature seen in the top right panel of Figure 2-3, which was described in the section on the elemental preference data.

The transformation T is interpreted as indicating that the effective value of a reinforcer is not proportional to the amount of food received but, rather, is related to that amount via a concave utility function $U(x) = x^{1/n}$. If the effective value of a food pellet of size x (in mg) is $x^{1/n}$, and the effective value, EV, of a sequence of reinforcers is a weighted sum of their individual effective values

$$\sum_{i=1}^{4} a_i (x_i)^{1/n} \quad ,$$

then the effective value of a substimulus (v_1, v_2, v_3, v_4) will be

$$EV = \sum_{i=1}^{4} a_i (v_i \cdot 20 \text{ mg})^{1/n}$$

$$= (20 \text{ mg})^{1/n} \sum_{i=1}^{4} a_i v_i \quad ,$$

whereas the effective value of the comparison schedule $(0 V_{mg} 00)$ will be

$$EV = 0 + a_2 (V_{mg})^{1/n} + 0 + 0$$

$$= a_2 (V_{mg})^{1/n} \quad .$$

The value V_{mg} at which indifference will be observed must be when

$$a_2 (V_{mg})^{1/n} = (20 \text{ mg})^{1/n} \sum_{i=1}^{4} a_i v_i \quad .$$

So that

$$V_{mg} = 20 \text{ mg} \cdot \left(\frac{\sum_{i=1}^{4} a_i v_i}{a_2} \right)^n \quad .$$

This is the regression equation, to be estimated next.

THE PREFERENCE DATA

The Models for the Decremental Contributions of Reinforcers to Effective Value

The models tested were the same as those in the perceived value case: multivariate, hyperbolic, exponential, and linear. (Note, however, one important difference: The value of n should be different from 1 in these regressions, because $T(V_{mg}) = V_{mg}^{1/n}$ is the transformation needed to get effective value, whereas n should equal 1 in the discrimination regressions, because the correct nonlinear transformation $z_{p(L)}$ has already been applied to the left side variable. An exponent greater than 1 has an interpretation in terms of a concave utility function for consumption of grain pellets in the case of the preference data, whereas it would have no such interpretation in the case of the discrimination data.) The

outcomes are shown in the bottom part of Table 2-5. The plots of the models for the derived points are shown in Figure 2-4. The derived points were found in the manner described for the discrimination data. Both the table and the figure show that the multivariate model accounts for the most variance in the data, the residual being the smallest. (This is necessarily so, as it is a more general model than the other three.) The hyperbolic accounts for the next most variance, doing slightly better than the exponential, with the linear doing quite a bit worse. Still the mean square errors were not different enough to produce significant F statistics between models. Table 2-5 lists one extra model tested for this procedure; it will be discussed in the next section.

The "forgetting" curves in Figure 2-4 were in the same order, and with similar spacing to the ones seen for the discrimination data. The variability of the preference data was much greater since the number of cases was smaller, so deviations might not be as meaningful.

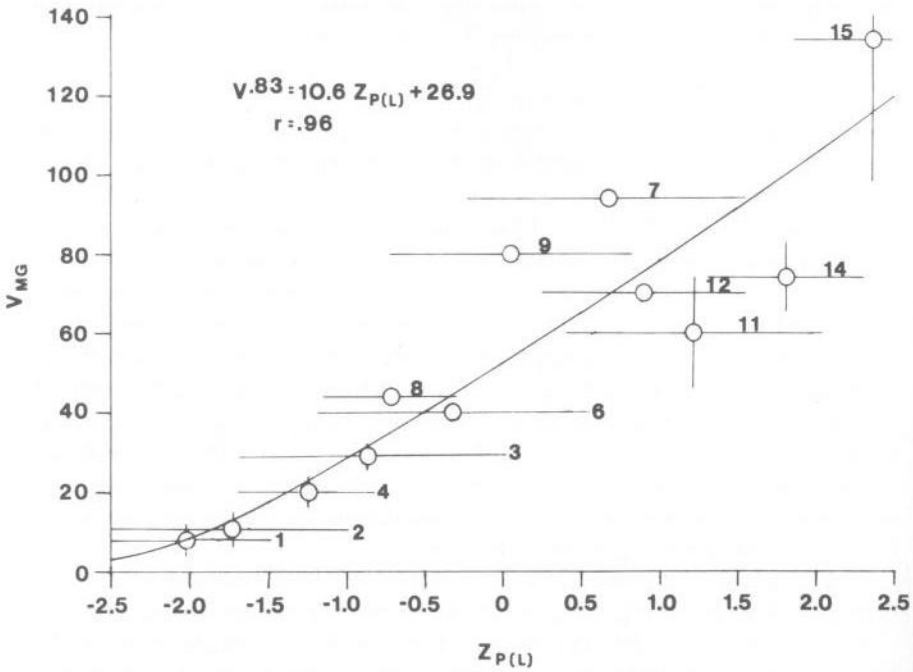
Perceived Value and Effective Value of a Substimulus Are a Linear Function of One Another: The Aggregation Process Must be Isomorphic. If control by reinforcers decreases as they recede from choice in the same manner in both the perceived value and effective value cases, then the function $a_i = f(c_i)$ should be the same in both cases. Not only the functional form—hyperbolic, exponential, or linear—should be the same, but the coefficients should be related.

The next-to-last model in Table 2-5 for the preference case used the values of a_i found in the 4-cycle discriminative multivariate model as estimates of the coefficients in the preference multivariate case. This procedure did quite well in accounting for variance and used only two parameters. The residual sum of squares was still less than that found for the linear model though more than the hyperbolic and exponential models. However, the mean square error (MS) was the smallest because the number of parameters estimated was only two. This shows that the coefficients from the discrimination case adequately account for the decremental weighting in the preference case.

This concludes the first way of showing that the perceived value of a sample from a schedule is directly related to the effective value of that same sample.

The second way of showing this is to plot directly the relation between the V_{mg} obtained in the preference experiment for a schedule sample and its perceived value $z_{p(L)}$ from the discrimination experiment. This is shown in Figure 2-5. The regression equation relating V_{mg} to $z_{p(L)}$ should yield a comparable exponent of 0.7 to the previous regression since $z_{p(L)}$ is a linear function of perceived value, as shown previously. Analytically this method of comparison is the same as the previous method but it is more traditional within psychology. The means of the values across birds for both the discrimination procedure and for the preference procedure at each of the 12 points were used in this second regression:

Figure 2-5. Titrated effective value V_{mg} is plotted against the perceived value $Z_{p(L)}$. The numbers refer to the substimulus numbers for either case. The lines indicate one standard deviation. Absence of a line in the vertical director means there is only one bird.



Transformed effective value = Linearly transformed perceived value

$$V_{\text{mg}}^{1/n} = \left(\sum_{i=1}^{4} b_i v_i \right)^m = a_0 + \sum_{i=1}^{4} a_i v_i = c_0 + c_1 z_{p(L)} \quad .$$

Logging both sides and performing the linear regression, the values obtained are

$$V_{\text{mg}}^{0.83} = 10.6\, z_{p(L)} + 26.9 \quad ,$$

or in a slightly different form:

$$V_{\text{mg}} = (10.6\, z_{p(L)} + 26.9)^{1.2} \quad .$$

The correlation $r = 0.9632$ between mean effective value ($V_{\text{mg}}^{.83}$) and mean perceived value $z_{p(L)}$ for each substimulus was reasonable.

Signal Detection Analysis of Molecular Matching and Maximizing: Is Molar Matching a Consequence of Molecular Processes Given Forgetting? The results of the discrimination experiment described in Commons (1979 and 1981) are presented as evidence of matching to relative expected payoff (programmed reinforcement) at the molar level. The relative $EP(L|S_{N_d})$ is the ratio of the expected payoff for a left-key-peck, $EP(L|S_{N_d})$, to the sum of the expected payoff for both key pecks:

$$\text{Rel } EP(L|S_{N_d}) = \frac{EP(L|S_{N_d})}{EP(L|S_{N_d}) + EP(R|S_{N_d})} \quad .$$

The expected payoff for a left-key peck in a given stimulus condition, $EP(L|S_{N_d})$, may be interpreted as the product of the conditional probability of reinforcement for a left-key peck in the given stimulus condition, $p(LC|S_{N_d})$ and its reinforcement utility U_L:

$$EP(L|S_{N_d}) = p(LC|S_{N_d}) \cdot U_L$$

$$EP(R|S_{N_d}) = p(RC|S_{N_d}) \cdot U_R$$

$$= [1 - p(LC|S_{N_d})] \cdot U_R \quad .$$

(We will allow for the possibility that the utility of reinforcement is not a linear function of the amount of food received.) Relative expected payoff retains the common notion of programmed relative amount of reinforcement (Catania, 1963). According to the matching law (Estes, 1957; Commons, 1979), the prob-

ability of pecking left for a substimulus of density N_d should be

$$p(L|S_{N_d}) = \text{Rel } EP(L|S_{N_d}) = \frac{EP(L|S_{N_d})}{EP(L|S_{N_d}) + EP(R|S_{N_d})} .$$

Now, for the 4-cycle T schedule used in Experiment 1, using Bayes' law:

$$\begin{aligned} EP(L|S_{N_d}) &= p(LC|S_{N_d}) \cdot U_L \\ &= \frac{p(S_{N_d}|LC)}{p(S_{N_d}|LC) + p(S_{N_d}|RC)} \cdot U_L \\ &= \frac{\binom{4}{N_d}\frac{3}{4}^{N_d}\ \frac{1}{4}^{(4-N_d)}}{\binom{4}{N_d}\frac{3}{4}^{N_d}\ \frac{1}{4}^{(4-N_d)} + \binom{4}{N_d}\frac{1}{4}^{N_d}\ \frac{3}{4}^{(4-N_d)}} \cdot U_L \\ &= \frac{\binom{4}{N_d}\ 3^{N_d}\ \left(\frac{1}{4}\right)^4}{\binom{4}{N_d}\ 3^{N_d}\ \left(\frac{1}{4}\right)^4 + \binom{4}{N_d}\ 3^{(4-N_d)}\left(\frac{1}{4}\right)^4} \cdot U_L \\ &= \frac{3^{N_d}}{3^{N_d} + 3^{(4-N_d)}} \cdot U_L \end{aligned}$$

Likewise,

$$\begin{aligned} EP(R|S_{N_d}) &= p(RC|S_{N_d}) \cdot U_R \\ &= \frac{p(S_{N_d}|RC)}{p(S_{N_d}|LC) + p(S_{N_d}|RC)} \cdot U_R \\ &= \frac{3^{(4-N_d)}}{3^{N_d} + 3^{(4-N_d)}} \cdot U_R . \end{aligned}$$

Hence, the matching law for programmed reinforcers would predict that

$$P(L|S_{N_d}) = \text{Rel } EP(L|S_{N_d})$$

$$= \frac{\dfrac{3^{N_d}}{3^{N_d}+3^{(4-N_d)}} \cdot U_L}{\dfrac{3^{N_d}}{3^{N_d}+3^{(4-N_d)}} \cdot U_L + \dfrac{3^{(4-N_d)}}{3^{N_d}+3^{(4-N_d)}} \cdot U_R}$$

$$= \frac{3^{N_d} \cdot U_L}{3^{N_d} \cdot U_L + 3^{(4-N_d)} \cdot U_R} .$$

The term U_L is the utility value of the reinforcer received for a correct left-key peck, and U_R is the utility value of the reinforcer received for a correct right-key peck. The following discussion uses $p(L)$ instead of $p(L|S_{N_d})$. The prediction of matching of expected payoff (also referred to as programmed reinforcement) can be simplified using a logistic transformation:

$$\frac{p(L)}{1-p(L)} = \frac{p(L)}{P(R)} = \frac{3^{N_d} \cdot U_L}{3^{(4-N_d)} \cdot U_R}$$

$$\log \frac{p(L)}{1-p(L)} = N_d \log 3 + \log U_L - (4-N_d) \log 3 - \log U_R$$

$$= \log \frac{U_L}{U_R} + (2N_d - 4) \log 3 . \qquad (2.1)$$

Hence molar matching predicts that the logistic transformation of $p(L)$ should be a linear function of the density N_d.

Figure 2-3 showed that if the average $p(L)$ value for all substimuli of a given density is plotted against N_d, the plot is of approximately the right shape. This suggests that the logistic transform of $p(L)$ is approximately a linear function of density of a sample at the molar level.

Unfortunately this theory does not have a systematic account of the great differences observed in the values of $p(L)$ for different substimuli of the same density. If the birds were really matching to relative expected payoff, given the information available to them, only the density of a given substimulus should matter. Instead, for substimuli of a given density, the birds are observed to peck left more frequently when the reinforcers occur later in the substimulus. In fact,

we have shown above that *molecular* level data for the zero-bias experiment ($U_L = U_R$) are well accounted for by a relation of the form:

$$z_{p(L)} = a_0 + \sum_{i=1}^{4} a_i v_i \quad , \tag{2.2}$$

where $v_i = 1$ for a substimulus with a reinforcer in cycle c_i, $v_i = 0$ for a substimulus with no reinforcer in that cycle, and the a_i are a decreasing sequence of positive weights given in Table 2-5. This account of the *molecular* level data implies that suitably transformed *molar* data should yield a linear plot, specifically:

$$z^*_{p(L|S_{N_d})} = \frac{N_d - 2}{4} \sum_{i=1}^{4} a_i \quad , \tag{2.3}$$

where $z^*_{p(L|S_{N_d})}$ represents the mean of the $z_{p(L)}$ values for all substimuli of a given density. (For derivation see Commons, 1981: 78-80.) The z value of $p(L)$ is found from the inverse probability of $p(L)$, also known as the *probit* transformation. Since this probit transformation of the data is not too different from the *logit* transformation used in Eq (2.1) for the range of values in question, a theory capable of predicting Equation (2.2) would also explain the general characteristic of the *molar* data that was interpreted as evidence of matching of expected value at the molar level.

Such a theory can in fact be constructed, on principles strictly compatible with utility maximization. The difference between this theory and the simple maximizing accounts rejected in Commons (1979) is that we assume that at the time of choice the bird can only remember the substimuli with error, the error being greater in the case of reinforcers more distant in time. (This was a maximization model suggested at the time but not developed.) The formal theory of utility maximizing choice under such conditions is well known as *signal detection theory* (Green and Swets, 1966; Egan, 1975). Since psychophysics experiments generally involve single stimuli, rather than sequences of stimuli of the sort we use, a full derivation of the implications of signal detection theory for an experiment of this sort is presented next.

Let y_i be the amount of reinforcement in a cycle c_i, as remembered by the bird at the time of choice. Assume that y_i is a normally distributed random variable with variance σ_i^2 and mean μ_i in the case that $v_i = 1$, and mean $-\mu_i$ and variance σ_i^2 in the case that $v_i = 0$. Assuming the two means to be additive inverses involves no loss of generality; only the ratio $d' = 2\mu_i / \sigma_i$ (to use the standard psychophysical index of discriminability of the reinforcer on cycle c_i) turns out to have observable significance. Four times the squared inverse of d', which is $\delta_i = \sigma_i^2 / \mu_i^2 = 4/(d')^2$, will be used instead of d' because it simplifies certain formulas. Then the distribution of y_i, in the case of the rich schedule, S_R : $v_i = 1$

with probability 3/4, $v_i = 0$ with probability 1/4, will have probability density function (PDF):

$$f(y_i) = \frac{3}{4} \frac{1}{\sqrt{2\pi}\,\sigma_i} \cdot e^{\frac{-(y_i-\mu_i)^2}{2\sigma_i^2}} + \frac{1}{4} \frac{1}{\sqrt{2\pi}\,\sigma_i} \cdot e^{\frac{-(y_i+\mu_i)^2}{2\sigma_i^2}} .$$

In the case of the lean schedule, $S_L : v_i = 1$ with probability 1/4, $v_i = 0$ with probability 3/4, it will have density function:

$$f(y_i) = \frac{1}{4} \frac{1}{\sqrt{2\pi}\,\sigma_i} \cdot e^{\frac{-(y_i-\mu_i)^2}{2\sigma_i^2}} + \frac{3}{4} \frac{1}{\sqrt{2\pi}\,\sigma_i} \cdot e^{\frac{-(y_i+\mu_i)^2}{2\sigma_i^2}} .$$

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 schedule sample, it will employ a likelihood ratio criterion. That is, it will peck left only when the ratio

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

Now,

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

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

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

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

and then expanding, yields

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

Then the term, $e^{\frac{-(y_i^2 + \mu_i^2)}{2\sigma_i^2}}$ may be factored out and cancelled top and bottom:

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

dividing top and bottom by $e^{\frac{-(y_i\mu_i)}{\sigma_i^2}}$ yields

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

The bird will peck left when this quantity is greater than 1. This is a nonlinear condition on the y_i values. But in the case that 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 approximations may be used:

$$\frac{3e^{\frac{(2y_i\mu_i)}{\sigma_i^2}} + 1}{3 + e^{\frac{(2y_i\mu_i)}{\sigma_i^2}}} \simeq \frac{3\left(1 + \frac{2y_i\mu_i}{\sigma_i^2} + \frac{2y_i^2\mu_i^2}{\sigma_i^4}\right) + 1}{3 + \left(1 + \frac{2y_i\mu_i}{\sigma_i^2} + \frac{2y_i^2\mu_i^2}{\sigma_i^4}\right)}$$

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

$$\simeq \left(1 + \frac{3}{2}\frac{y_i \mu_i}{\sigma_i^2} + \frac{3}{2}\frac{y_i^2 \mu_i^2}{\sigma_i^4}\right)\left(1 - \frac{1}{2}\frac{y_i \mu_i}{\sigma_i^2} - \frac{1}{4}\frac{y_i^2 \mu_i^2}{\sigma_i^4}\right)$$

$$\simeq 1 + \frac{y_i \mu_i}{\sigma_i^2} + \frac{1}{2}\frac{y_i^2 \mu_i^2}{\sigma_i^4} \simeq e^{y_i \mu_i / \sigma_i^2} .$$

So the condition for the bird to peck left is

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

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

Taking the log of both sides gives

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

The approximations are quite robust. 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 $EP(L)/EP(R) \simeq 1$; that is, where

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

which in the case of only moderate bias, will mean y_i values such that none of the $(y_i \mu_i)/\sigma_i^2 = y_i/\delta \mu_i$ are very large.

We have found that the bird will peck left when

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

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

$$M = \sum_{i=1}^{4} \frac{(2v_i - 1)}{\delta_i}$$

and variance

$$V = \sum_{i=1}^{4} \frac{1}{\delta_i}$$

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

$$z_{p(L)} = \frac{M - \ln (U_R / U_L)}{V^{1/2}} \tag{2.4}$$

so that

$$z_{p(L)} = \sum_{i=1}^{4} a_i v_i + b \quad , \tag{2.5}$$

where

$$a_i = \frac{2}{\delta_i \left[\sum_{j=1}^{4} \frac{1}{\delta_j} \right]^{1/2}} \tag{2.6}$$

and

$$b = \frac{\ln (U_L/U_R) - \left[\sum_{j=1}^{4} 1/\delta_j \right]}{\left[\sum_{j=1}^{4} 1/\delta_j \right]^{1/2}} \quad .$$

We thus obtain the linear form of Eq. (2.2) for the zero bias case ($U_L = U_R$). In addition to predicting the linear form (2.2), the theory predicts that one should find

$$\sum_{j=1}^{4} a_j = 2 \left[\sum_{j=1}^{4} \frac{1}{\delta_j} \right]^{1/2} = -2a_0 \quad .$$

This relation is approximately satisfied by the values given in Table 2-4,

$$-2a_0 = 5.084 \quad , \quad \sum_{j=1}^{4} a_j = 4.583 \quad .$$

The slight deviation probably indicates that the variances of the two normal distributions are not equal, as assumed previously, when $v_i = 1$ and 0, respectively.

The model can easily be generalized although the generalization will not be discussed here. We also obtain the result, verified by Commons (1979), that the effect of bias should be a vertical shift of the plot of suitably transformed response probability against density. (This molar result, of course, is also a prediction of the matching eq. (2.1).)

The Decrementing of Value: The Additive Noise Model of Forgetting Functions

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 Eq. (2.6):

$$\delta_i = \frac{4}{a_i \left(\sum_{j=1}^{4} a_j \right)} \quad . \tag{2.7}$$

For the data to be consistent with a model of maximizing with imperfect discrimination, then, the linear relation of Eq. (2.2) must be observed, and the a_i obtained must be such that

$$a_1 > a_2 > a_3 > a_4 > 0 \ .$$

The values reported in Table 2-5 generally satisfy this condition. More precise predictions as to the relation that should exist among the a_i depend upon a particular model of forgetting. It makes more sense to formulate a forgetting model in terms of the δ_i rather than the a_i since it is the δ_i that have a simple interpretation in terms of discriminability.

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_o (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_o 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 4-cycle experiment), can be written as

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

Therefore,

$$E(Y_i) = E(Y_o)$$

$$\text{var}\ (Y_i) = \text{var}\ (Y_o) + i \cdot \text{var}(e)$$

and

$$\delta_i = \frac{\text{var}\ (Y_i)}{E(Y_i)^2} = \frac{\text{var}\ (Y_o)}{E(Y_o)^2} + i \cdot \frac{\text{var}\ (e)}{E(Y_o)^2} \quad .$$

The discriminability measure δ_i, as a function of cycle number i, is therefore predicted to be a linear function:

$$\delta_i = \delta_o + i\gamma \quad .$$

Substituting this expression into (2.6) yields

$$a_i = \frac{c}{d+i} \tag{2.8}$$

where

$$c = \frac{2}{\gamma}\left[\sum_{j=1}^{4} (\delta_o + j\gamma)^{-1}\right]^{-\frac{1}{2}}$$

and

$$d = \frac{\delta_o}{\gamma} \quad .$$

This is a hyperbolic forgetting function of the sort discussed previously and estimated as line 3 of Table 2-4. As we have seen, the data are consistent with a forgetting function of this form; the 6-cycle data in particular indicate a function of this form. It can be shown that Eq. (2.8) holds for the general N-cycle case, so that the same hyperbolic function is observed regardless of the number of cycles.

Furthermore, quite general predictions are generated regarding the relation that should exist between the a_i values found for experiments involving different numbers of cycles, independent of the particular forgetting model used. If it is assumed that the discriminability of a reinforcer occurring, say, 2 cycles before choice should remain the same no matter how many cycles preceded it, then precise predictions are obtained regarding the a_i values that should be found for, say a 6-cycle experiment, and those found for the 4-cycle experiment. It is not that the last 4 a_i should be exactly the same in both cases; rather, the a_i should be such that when inserted into Eq. (2.7), the last 4 δ_i obtained are exactly the same in both cases.

In particular this theory predicts that the linear plots of suitably transformed response probability against density will have less steep a slope the greater the number of cycles in the experiment. The molar matching Eq. (2.1) generalizes to

$$\log \frac{p(L)}{1-p(L)} = \log \frac{U_L}{U_R} + (2N_d - N) \log 3 \quad , \tag{2.1'}$$

where N is the number of cycles. This implies that the plot should have the same slope no matter what the number of cycles, as shown in the top left panel of Figure 2-6. The linear plot predicted by the maximizing theory on the other hand, generalizes to

$$z^*_{p(L|S_{N_d})} = \frac{2N_d - N}{2N} \sum_{i=1}^{N} a_i \quad , \tag{2.3'}$$

where Eqs. (2-4)-(2-7) remain the same except that all summations should be from 1 to N. From (2.6) we see that the slope of (2.3′) is

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

When another cycle is added to the experiment, a new term is added to the sum. Since the δ_i for the new cycle must be larger than any of the others, the average value

$$\frac{1}{N} \sum_{i=1}^{N} \delta_i^{-1}$$

decreases. Since this term in the numerator decreases and the N in the denominator increases, the slope of (2.3′) decreases, as shown in the top right panel of Figure 2-6 for the 4- and 6-cycle cases. Hence the predictions of the maximizing theory and the matching theory can be distinguished even for *molar* data. The *molar* data plots presented in Figure 2-3 support the maximizing theory, since the 6-cycle plot has less steep a slope. Equation (2.1′) also indicates that the matching of expected payoff theory predicts that the degree of vertical shift associated with a given bias condition should be independent of the number of cycles. Equation (2.6) on the other hand, indicates that the maximizing of expected payoff theory predicts that the degree of vertical shift associated with a given bias condition should be similar the larger the number of cycles. This is another difference at the level of *molar* data alone.

In the case of the additive noise model (hyperbolic forgetting function), the slope of (2.3′) is

$$\frac{2}{N} \left(\sum_{i=1}^{N} \delta_i^{-1} \right)^{1/2} = \frac{2}{N} \left(\sum_{j=1}^{N} (\delta_0 + j\gamma)^{-1} \right)^{1/2} .$$

Figure 2-6. Theoretical plots (*top right*) of the value $V_{S_{N_d}}$ (on the y axis) that is the logit

$$p(L|S_{N_d}) = \log \frac{p(L|S_{N_d})}{1 - p(L|S_{N_d})} = \log p(L|S_{N_d}) / p(R|S_{N_d})$$

on the top left panel and probit $p(L) = z_{p(L|S_{N_d})}$. There are two x axes; the one on the top is for the number of reinforcers in a substimulus in the 4-cycle experiment and the one on the bottom is for the 6-cycle experiment. The middle panel is also a theoretical plot of $p(L)$ versus either the z transform of $p(L)$ or the phi transform of $p(L)$. Note that the slope in the center portion is the same but that the knee at A is sharper than at B. Also the same is true for the ceiling knee although this is not marked. In the theoretical plot of $p(L)$ versus the predicted transformed value, $\hat{z}_{p(L)}$ (*bottom left*), if the straight line is more descriptive, local maximizing is supported, whereas if the curved line is more descriptive, local matching is supported. In two theoretical plots of the mean of the residuals $z_{p(L)} - \hat{z}_{p(L)}$ versus $\hat{z}_{p(L)}$ (*bottom middle*), the dashed line is for local matching and the x axis represents local maximizing. The plot of the actual residuals against $\hat{z}_{p(L)}$ is shown bottom right.

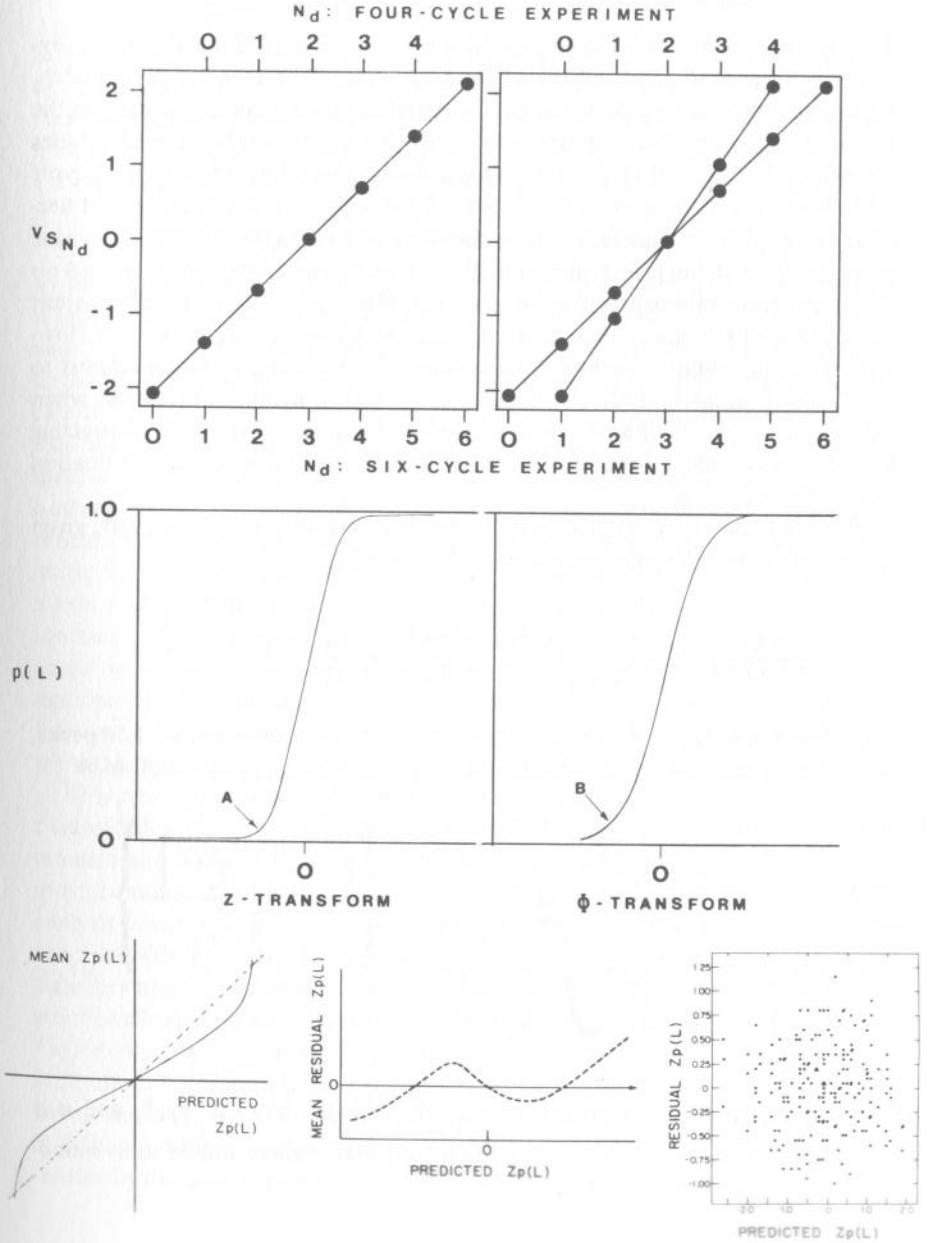
This implies that given the slopes for any two values of N, the values of δ_0 and γ may be calculated, and from them predictions obtained for the slopes for all other values of N. Thus it is possible to test the additive noise model at the level of *molar* data alone, even though there is no obvious forgetting function at that level.

The most interesting test of the additive noise model, of course, compares the values for δ_0 and γ implied by *molecular* level data (for any given number of cycles), with the values implied by comparative *molar* level data between experiments with different numbers of cycles. The 6-cycle data, as analyzed, indicates a value of $d \simeq 7.4$ (where d is presented in equation (2.8)), which would imply $\delta_0 \simeq 0.48$, and hence that:

$$\frac{\text{slope } (N = 6)}{\text{slope } (N = 4)} \simeq \frac{4}{6} \cdot \left[\frac{\frac{1}{1.4} + \frac{1}{2.4} + \frac{1}{3.4} + \frac{1}{4.4} + \frac{1}{5.4} + \frac{1}{6.4}}{\frac{1}{1.4} + \frac{1}{2.4} + \frac{1}{3.4} + \frac{1}{4.4}} \right]^{1/2} \simeq 0.7 ,$$

which is close to what is observed.

Figure 2-6


N_d: FOUR-CYCLE EXPERIMENT
$V_{S_{N_d}}$
N_d: SIX-CYCLE EXPERIMENT
p(L)
A
B
Z-TRANSFORM
Φ-TRANSFORM
MEAN Zp(L)
PREDICTED Zp(L)
MEAN RESIDUAL Zp(L)
PREDICTED Zp(L)
RESIDUAL Zp(L)
PREDICTED Zp(L)

Predictions of Local Matching Contrasted with Predictions of Local Maximizing

To the claim that a maximizing model with imperfect discriminability can explain both the *molecular* and the *molar* data better than does the molar matching law it might be objected that such a matching model could also explain the fact that different substimuli of the same density result in different probabilities of pecking left. In reply it can be observed that the matching of expected payoff hypothesis was introduced in Commons (1979) only because it was thought necessary to explain the shape of the *molar* plots of $p(L)$ against density; we have now shown that imperfect discrimination is sufficient to explain that, and no deviations from rationality need be assumed. Many other examples of apparent matching might similarly be explained as due to imperfect discrimination. However, if the matching hypothesis is considered of sufficient theoretical interest to justify contrasting it with the hypothesis of utility maximization even when thre are no gross qualitative results found to be inconsistent with maximizing behavior, then predictions of a matching model with imperfect discrimination may be easily derived.

It would then be predicted that the bird's probability of pecking left, given its recollection of the substimulus would be given by

$$\frac{p(L)}{1-p(L)} = \frac{EP(L\,|\,y_1, y_2, y_3, y_4)}{EP(R\,|\,y_1, y_2, y_3, y_4)} = \frac{U_L}{U_R} \cdot e^{\sum_{i=1}^{4} \frac{\mu_i y_i}{\sigma_i^2}} ,$$

using the same approximation as before. The total proportion of left pecks, when the actually occurring substimulus is v_1, v_2, v_3, v_4, will therefore be

$$p(L) = \frac{1}{\left[2\pi \sum_{i=1}^{4} \delta_i^{-1}\right]^{1/2}} \int_{-\infty}^{\infty} \frac{e^z}{\frac{U_R}{U_L} + e^z} \, e^{\left\{ \frac{\left[z - \sum_{i=1}^{4} \frac{(2v_i - 1)}{\delta_i} \right]^2}{2 \sum_{i=1}^{4} \delta_i^{-1}} \right\}} dz .$$

This expression depends on the v_i only through the sum $\sum_{i=1}^{4} v_i/\delta_i$ and is a monotonically increasing function of that sum. Hence there will be some monotonic transformation Φ of $p(L)$ such that

$$\Phi\, p(L) = \sum_{i=1}^{4} a_i v_i , \qquad (2.9)$$

where the relative sizes of the a_i will again have the interpretation,

$$a_i / a_j = \delta_j / \delta_i \quad .$$

In particular, this theory also predicts that if $\delta_4 > \delta_3 > \delta_2 > \delta_1$, then $a_1 > a_2 > a_3 > a_4$. The definitive comparison of the local matching to the local maximizing hypothesis, then, will require careful examination of the transformation of the $p(L)$ data required to get a linear relation of the form (2.9). If a z transformation is correct, then local maximizing is indicated. Local maximizing theory would predict the relation of $z_{p(L)}$ to $a_0 + \sum_{i=1}^{4} a_i v_i = (\hat{z}_{p(L)})$ to be the straight line shown in the bottom left panel of Figure 2-6. Whereas if the correct transformation is a significantly gentler one (like the logit or arcsine or the one that is shown by the curve in the bottom left of Figure 2-6, deviation from maximizing behavior in the direction of local matching is indicated. Therefore local matching will be indicated by consistently positive residuals for the data points with large positive values for $\hat{z}_{p(L)}$, and consistently ngative residuals for the data points with large negative values for $\hat{z}_{p(L)}$. The bottom middle panel of Figure 2-6 shows another version of the same theoretical plot. If local matching were true, the mean of the residuals at each predicted $\hat{z}_{p(L)}$ value would be the curve. If the means of the residuals fell on the x axis, local maximizing would be true, since the x axis represents the fact that the mean of the residuals should not deviate from 0 as a function of $\hat{z}_{p(L)}$. The actual residuals, plotted against the predicted $\hat{z}_{p(L)}$, are shown in the bottom right panel of this figure. There is no evidence that the mean of the residuals follows the local matching curve. The data are consistent with local maximizing where the variance changes as a function of $\hat{z}_{p(L)}$.

This test could have been complicated by the fact that the derivation of the z transform in the case of maximizing theory relies upon the assumption that the remembered values of reinforcement y_i are normally distributed. This is the most common assumption in the applications of signal detection theory; the analysis could equally well be carried out for many other families of distribution functions (Egan, 1975). If the proper distribution is not normal, then the probit transformation should not yield a perfect linear regression even in the case of strict local maximizing behavior. Our derivation of Eq. (2.5) also depends on a Taylor series approximation, so that even if the assumption of normality is correct, one should not expect a perfect linear regression. Hence distinguishing between the fully general maximizing model and the fully general matching model is difficult. Still it is important to note that this analysis has revealed *no* regularities in the data that are not consistent with a generalized maximizing model.

Ethological Generalizations

The present situations bear on ethological interpretations of choice. Different patches in the environment may have changing amounts of reinforcement. The spatial distribution of reinforcers (prey) is converted into a temporal distribution of response-produced reinforcement (a reinforcement schedule) as the organism forages. The amount of prey in a patch may vary systematically from one patch to another by being richer or leaner so that the rate of obtaining reinforcement is either higher or lower. The prey density may also vary randomly so that one subpatch might be leaner than the average of another patch but the patch from which it comes may be richer overall. How does a forager in the nonhuman case and a decisionmaker in the human case combine those reinforcers within a subpatch (schedule sample)?

In the foraging situation the uneven distribution of prey may be decomposed into nonsystematic variation (noise) and systematic changes in prey density (signal) over an area that takes a period of time to obtain. Two sets of decisions might be made. First, the organism, from examining the sample, may forecast how dense the patch is. This is the discriminated or perceived value of the sample. This discriminative information can be used to control or shift behavior by serving as a cue to shift or stay. Second, the organism's decision to continue foraging may be reinforced with respect to the other experience yields. This is the effective value of the sample from the patch.

Hence there are two sets of behavior. One set consists of the choice to stay or to leave (shift) and may be controlled by the perceived value of the reinforcement sample. The second set obtains the reinforcers and is under the direct control of the bundle of reinforcers obtained in the schedule sample. Both economic (e.g. Pareto, 1906) and ethological (e.g., Krebs, Kacelnik, and Taylor, 1978; Lea, Chapter 7; and Kamil, Peters and Lindstrom, Chapter 8); theories have generally assumed that organisms maximize expected payoff in a choice situation. However, in psychology (e.g., Estes, 1957), interpretations of data on choice have tended to support the notion that organisms match their choice probabilities to the relative rate of the programmed outcome. At first glance these two sets of theories seem to be incompatible. In the type of choice situation studied here, however, momentary maximizing at the molecular level may result in overall molar matching. The underlying processes that establish the utility of the schedule sample may involve momentary maximizing when forgetting and imperfect discriminability of the sample exist.

CONCLUSION

Since the aggregation functions in the discrimination case and the preference case have been shown to be equivalent, a similar linear noise model should

account for both. Models of the preference case and how local maximizing and an additive noise model might work there are left for a later exposition for a number of reasons.

First, it is necessary to have an analysis of acquisition data that reflects what is going on as response strength on each preference key shifts. This will be discussed by Commons, Woodford, Ducheny, and Peck (forthcoming). Second, it would be useful to have a better idea what the a_i weights look like for reinforcers occurring further from choice. A discrimination experiment with 6 to 18 cycles (Cabral, Woodford, and Commons, 1982) will allow for a more accurate determination of which model of perceived value produces the correct weighting scheme.

To determine better the correct weighting scheme in the preference situation, a new procedure must be used. Since the data for each point in the preference experiment (a concurrent chain titration schedule) take almost a year to collect, obtaining a sufficient number of points using that procedure with twelve cycles would be prohibitive. A new discrimination paradigm with the schedule samples as a consequence for correctly discriminating two similar stimuli would more rapidly establish the samples' values. By using signal detection analysis those values can be calculated from the bias the samples create.

It is likely that a similar principle to that of momentary maximizing of choice payoff will be true for effective utility of a schedule sample. The macro, molar, and molecular results for,the preference and discrimination experiments were homologous. The process that creates perceived value and effective value of a schedule sample should be very similar. If this is true, however, it would mean that behavior in the preference experiment is largely determined by discriminative properties of the substimuli. Problems in the discriminability of substimuli should manifest themselves most clearly in the rate at which behavior adjusts when the substimulus is changed. Hence in the case of the preference experiment, the best way to determine the type of noise process involved is by looking at acquisition, which will be the subject of a future analysis. Presently, undermatching of obtained payoff has been shown to be consequence of forgetting and of imperfect sensitivity to schedule differences. The degree of undermatching is a function of the amount of forgetting and of imperfect sensitivity to schedule differences. Local maximizing therefore will account for undermatching and obtained reinforcement that varies in degree from none to great. Obtained reinforcement matching accounts for obtained reinforcement maximizing; local maximizing and the additive noise processes should account for the traditional matching and maximizing of obtained payoff phenomena and predict when these accounts will hold.

Since aggregate obtained matching and aggregate obtained maximizing accounts predict the same local behavior in a concurrent VR VR situation, it would be good to have an experimental situation in which obtained matching and maximizing make different predictions. Local maximizing theory should be

able to predict that, in a VR VI concurrent situation, matching to obtained reinforcement at the molar level is supported. The way to go about this is to modify the discrimination procedure described previously in this chapter. Remember that the stimuli to be discriminated overlap so that the concurrent FR-1, extinction programmed under any given stimulus condition appears as a rich VR versus a very lean VR for making a left versus right choice. The discriminability, or the lack thereof, of the stimuli indicating which side-key peck has the FR-1 program on it, controls the degree to which that FR-1 schedule appears as a VR schedule. With perfect discriminability one key would appear to have a FR-1 and the other key extinction. With no discriminability both keys would appear to have equal VR schedules programmed on them.

In a proposed discrimination experiment with imperfectly discriminable stimuli, each correct left choice is considered one behavioral unit. Left correct choices would be reinforced on a VI schedule. The other behavioral unit, a right correct choice, would be reinforced on a VR schedule that would yield approximately the same obtained rate of reinforcement as the left key. Suppose that each trial is 18 seconds long, 12 seconds for the stimulus period and 6 seconds for the choice period. The average of the VI intervals would be 72 seconds. The first left correct after the interval period (which are multiples of 18 seconds timed-out would be reinforced. The right corrects would be paid off on a VR schedule. For instance, the fifth right correct would be reinforced. The maximum programmed probability of reinforcement per trial for making a left choice on the VI would be 0.2, that is, the first left correct after four complete stimulus periods. For the VR it would be 0.2 also. The difference between the two is the VI does not require on the average four previous correct right choices, only that they had been programmed, whereas the VR requires four previous correct right choices. This would favor choosing the left VI key if the obtained reinforcement matching law described by Herrnstein (1970) is more correct than obtained reinforcement maximizing. The undermatching predicted by our local reinforcement maximization would produce undermatching, more indifference between the two alternatives than even predicted by obtained reinforcement matching. Obtained reinforcement maximization would favor almost exclusive preference for the VR key since only one choice out of five on the average has to go to the VI key to peck up at least half the reinforcers programmed and saved on it. It is hoped that the complexities of the situation will not make it impossible to extend the local programmed reinforcement maximizing model to this case.

REFERENCES

Baum, W.M., and H.C. Rachlin. 1969. Choice as time allocation. *Journal of the Experimental Analysis of Behavior 12*: 861–74.

Cabral, B.D.; M. Woodford; and M.L. Commons. 1982. Local maximizing versus local matching; How perceived values of reinforcement schedule samples that vary in length bear on those proposed mechanisms underlying matching found at the aggregate level. Presented at the Eastern Psychological Association.

Catania, A.C. 1963. Concurrent performances: A baseline for the study of reinforcement magnitude. *Journal of the Experimental Analysis of Behavior 6*: 299–301.

Chung S., and R.J. Herrnstein. 1967. Choice and delay of reinforcement. *Journal of the Experimental Analysis of Behavior 10*: 67–74.

Commons, M.L. 1979. Decision rules and signal detectability in a reinforcement density discrimination. *Journal of the Experimental Analysis of Behavior 32*: 101–20.

____. 1981. How reinforcement density is discriminated and scaled. In M.L. Commons and J.A. Nevin, eds., *Quantitative Analyses of Behavior*, vol. 1, *Discriminative Properties of Reinforcement Schedules*, Cambridge, Mass.: Ballinger.

Commons, M.L.; M. Woodford; J.R. Ducheny; and J.R. Peck. Forthcoming. The acquisition of performance during shifts between concurrent schedule terminal links. In M.L. Commons, R.J. Herrnstein, and A.R. Wagner, eds., *Quantitative Analyses of Behavior*, vol. 3, *Acquisition*, Cambridge, Mass.: Ballinger.

Egan, J.P. 1975. *Signal Detection Theory and ROC Analysis.* New York: Academic Press.

Estes, W.K. 1957. Theory of learning with constant, variable, or contingent probabilities of reinforcement. *Psychometrika 22*: 113–32.

Green, D.M., and J.A. Swets. 1966. *Signal Detection Theory and Psychophysics.* New York: John Wiley.

Herrnstein, R.J. 1970. On the law of effect. *Journal of the Experimental Analysis of Behavior 13*: 243–66.

Herrnstein, R.J., and G.M. Heyman. 1979. Is matching compatible with reinforcement maximization on concurrent variable interval variable ratio? *Journal of the Experimental Analysis of Behavior 31*: 209–24.

Herrnstein, R.J., and W. Vaughan. 1980. Melioration and behavioral allocation. In J.E.R. Staddon, ed., *Limits to Action: The Allocation of Individual Behavior.* New York: Academic Press.

Krebs, J.R.; A. Kacelnik; and P.J. Taylor. 1978. Test of optimal sampling by foraging great tits. In *Nature 275*: 27–31.

Mandell, C. 1981. A psychophysical analysis of time-based schedules of reinforcement. In M.L. Commons and J.A. Nevins, eds., *Quantitative Analyses of Behavior*, vol. 1, *Discriminative Properties of Reinforcement Schedules*, Cambridge, Mass.: Ballinger.

Mazur, J.E. 1981. Optimization theory fails to predict performance of pigeons in a two-response situation. *Science 214*: 823–25.

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

Neimark, E.D., and W.K. Estes. 1967. *Stimulus Sampling Theory.* San Francisco: Holden-Day.

Nevin, J.A. 1969. Signal detection theory and operant behavior. Review of D.M. Green and J.A. Swets, *Signal Detection Theory and Psychophysics. Journal of the Experimental Analysis of Behavior 12*: 475–80.

Pareto, V. 1906. Ophelimity in nonclosed cycles. *Giornale degli economisti* (2) 33 November. Also translated by A.P. Kirman in J.S. Chipman, L. Hurwiez, M.L. Richter, and Sonnenschein, eds., *Preferences, Utility, and Demand*, Harcourt, Brace and Jovanovich, 1971.

Rachlin, H.C., and L. Green. 1972. Commitment, choice and self-control. *Journal of the Experimental Analysis of Behavior 17*: 15–22.

Rachlin, H.; Battalio, R.; Kagel, J.; and L. Green. 1981. Maximization theory in behavioral psychology. *Behavioral and Brain Science 4*: 371–417.

Schoenfeld, W.N., and B.K. Cole. 1972. *Stimulus Schedules: The T–V Systems.* New York: Harper & Row.

Schoenfeld, W.N.; W.W. Cummings; and E. Hearst. 1956. On the classification of reinforcement schedules. *Proceedings of the National Academy of Sciences 42*: 563–70.

Weissman, A. 1961. Impairment of performance when a discriminative stimulus is correlated with a reinforcement contingency. *Journal of the Experimental Analysis of Behavior 4*: 365–69.

Wickelgren, W.A. 1974. Strength/resistance theory of the dynamics of memory atorage. In D.H. Krantz, R.C. Atkinson, R.D. Luce, and P. Suppes, eds., *Contemporary Developments in Mathematical Psychology: Learning, Memory, and Thinking.* San Francisco: Freeman.