3 HOW REINFORCEMENT DENSITY IS DISCRIMINATED AND SCALED

Michael L. Commons

Understanding how reinforcement schedules are discriminated and scaled should help explain behavior occurring in two different situations. The first of these is where some property of a schedule sample acts as a cue for future behavior, either in the laboratory or in the field (Kamil, Peters, and Lindstrom, 1981; Williams, 1981). As one important example, the effective value of a series of reinforcers may determine upcoming decisions of the subject—for instance, in a foraging situation where the relative density of reinforcement for foraging in a patch controls the choice to stay in that patch or to shift to another. In the second situation, understanding the perceived or scaled value of schedule samples should be useful in explaining the strengthening effects of reinforcement schedules on operant behavior. This is likely because of the high degree of correlation between

---

Portions of this chapter were reported at 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. A number of people have contributed to the analysis. They include, in alphabetical order, Michael Cohen, John R. Ducheny, Richard J. Herrnstein, Gene Heyman, Duncan Luce, John A. Nevin, Joel R. Peck, William Vaughn, and Michael Woodford. Rose P. Meegan and John R. Ducheny wrote the computer program for the analysis. Computer facilities at Boston University and Northeastern University were used. The organization of the complex material is the work of Patrice M. Miller, who edited the text. The line of research leading up to the present research was developed under the guidance of John Anthony Nevin. The analysis was carried out with the support of Richard J. Herrnstein. Research support was also provided by the Dare Association, Inc.

perceived value and choice, as shown by Commons and Ducheny (1981). For example, in chained schedules of reinforcement, the perceived value of the reinforcers in the terminal link may be predictive of behavior in the initial link, whose only reinforcement is the occurrence of the terminal link.

The primary purpose of the present studies is to explore the processes by which pigeons discriminate and scale reinforcement density. The subjects obtained occasional reinforcers on schedule samples for center-key pecks (R_C). These schedule samples then served as stimuli to be discriminated. After obtaining the sample, a pigeon indicated whether the sample was from a rich or a lean schedule by a peck to either the left or the right key. In previous work using this method, Commons (1979) found that on the average, pigeons matched probability of a choice, indicating which sample had occurred, to the relative expected payoff for making that choice. They did not maximize payoff by exclusively choosing the appropriate response alternative, as an ideal observer would. However, departures from ideal observer performance may depend on the temporal pattern of reinforcers within the sample. Samples could have reinforcement occurring during any of four sequentially presented 3-second cycles. Therefore, a single reinforcer might occur temporally just before choice or temporally remote from choice. A sample with a single reinforcer occurring right before choice produced more left-key peck choices (the peck indicating that the sample came from the rich schedule) than the sample with the single reinforcer 9 seconds earlier. Choices after samples when the time of reinforcement was either just before or remote did not follow the matching law, whereas the average decision across samples did. Similarly, a sample with a single omitted reinforcer (having three rather than four possible reinforcers) placed immediately before choice produced fewer left-key pecks than a sample with the single omitted reinforcer placed 9 seconds before choice. The change was less for the single missing reinforcer, suggesting that the presence of reinforcement was more salient than its absence.

These previous findings suggested that forgetting might account for the measured discriminability and perceived value of the samples from the two schedules, where forgetting is defined as a decrement in control by a prior event over a future choice as a function of time. The question now is, How might one conceive of a detection and scaling mechanism that is reasonably sensitive to average reinforcement density, but is sent awry by changes, due to forgetting, in the temporal distribution of events within a sample?

To analyze the detection and scaling process, the effect of changing the amount of time between reinforcement opportunities in a schedule sample and the length of the sample were studied in order to learn how decision rules and isosensitivity functions (see Nevin, Chapter 1 for definitions of these two terms) depend on forgetting within a sample in a density of reinforcement discrimination situation. Three experimental manipulations were used to answer these questions. In two of these, cycle length in the four-cycle sample was changed. In the third, the number of cycles was changed, while base (standard) cycle length was kept constant. Cycle length was changed in two ways—either to produce a change to a new cycle length for an extended number of sessions (stable) or to produce a momentary change on a small set of trials within a session (probe).

It is proposed that the determination of perceived density is at least a two-step process. First, summative trace decay theory (Wickelgren, 1974) should account for the relative contribution of each reinforcer within a schedule sample. Summative trace decay theory is modified here and is called the relative-time-weighted sum theory, suggesting that the value of each reinforcer is weighted by how it is discounted, how well it is remembered (Wickelgren, 1974), and that the weighted values are summed. It would predict that as cycle length is increased, whether in a stable or a temporary fashion, the perceived value contributed by each reinforcer should decrease. Second, a set of theories, here referred to as base rate or base time theory, suggests that steady versus momentary changes in cycle length have different effects on perceived density of a sample. One version suggests that a time base and time window are established and that perceived density of the sample reinforcement rate is found relative to those bases. For instance, doubling the cycle length on a probe trial halves the actual rate of reinforcement, which may cause the perceived density to be scaled as half its value on standard or base cycle length trials. This also would leave half the possible events occurring outside the base time window, thus lowering the perceived value further. The effects of adding more events by adding more cycles required explicit examination.

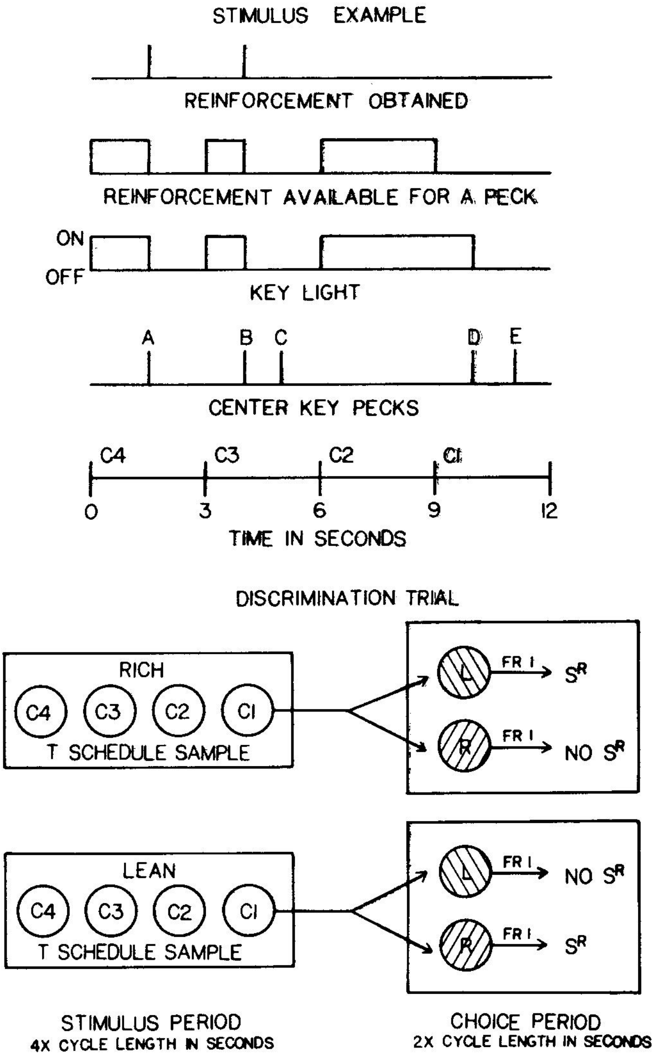
By increasing the number of cycles in a sample while keeping maximally obtainable discriminability constant and comparing the results to those obtained with an equivalently long schedule sample, it is possible to examine the relative contribution of these two sources of discounting information within the sample. First, discounting in the form of forgetting occurs because time between reinforcement opportunities increases and the corresponding number of unprogrammed events increases. Second, as the number of cycles presented within the stimuli to be discriminated increases, the information-processing capacity of the organism is taxed, interference results, and forgetting occurs (D'Amato, 1973; Grant and Roberts, 1973). In either case, both signal detection measures of discriminability and sensitivity (d') and measures of bias ($p(L)$) indicating that the rich schedule has been presented, should decrease, but for different reasons. In Experiment I the effect of changing cycle length is explored.

EXPERIMENT I: PROCEDURE

Birds were run in one 256-trial session per day. Trials consisted of a stimulus period followed by a choice period, as shown in Figure 3-1. The present task is similar to one in which the subject identifies whether a randomly chosen urn is the rich one that has three reinforcer balls and one nonreinforcer ball or whether the urn is the lean one that has one reinforcer ball and three nonreinforcer balls. On each trial there are four draws from that trial's urn. 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.

Here, the stimuli to be discriminated were rich, $p(S^{R+}|R_C) = 0.75$, and lean, $p(S^{R+}|R_C) = 0.25$, schedules for center-key pecks, R_C presented as modified T schedules (Commons, 1979; Schoenfeld and Cole, 1972; Schoenfeld, Cumming, and Hearst, 1956; Weissman, 1961). On each trial, a sample from one of the two T schedules was presented during the stimulus period. Each one of these sixteen substimulus samples consisted of four equal duration cycles. On each cycle, a center-key peck was ($v_i = 1$) or was not ($v_i = 0$) reinforced. The cycles, c_i, were numbered so that *C1* is the cycle at the end of the stimulus period and right before the choice period and *C4* is at the beginning of the stimulus period and the furthest from the choice. An unlikely example of what might happen during a presentation of a 1110 substimulus sample is shown in the upper portion of Figure 3-1. As is shown, each cycle began with the illumination of

Figure 3–1. The top portion shows a state diagram for a sample from a 3-second cycle schedule. It illustrates what may happen if center-key pecks occur and do not occur when reinforcement has been programmed or not. It is for illustrative purposes only and would be very unusual because there is a cycle without a peck occurring in it. The bottom portion shows the contingencies during an entire trial. The stimulus period contains substimuli of the form shown in the top portion. The choice period immediately follows the stimulus period. There is no intertrial interval.



the center key. The first center-key peck darkened the key and was reinforced with the same probability as the rest of the cycles on that trial, $p(S^{R+}|R_C)$ being either 0.75 or 0.25. No other center-key pecks during that cycle were reinforced, although they occurred.

At the onset of the choice period the side keys were illuminated and the center key stayed dark or was darkened in those rare cases where no key peck occurred in the last cycle of the substimulus. The choice period duration was always twice the standard or base length cycle, as shown in the lower portion of Figure 3-1. The first side-key peck, whether correct or not, darkened both keys, and no further pecks were counted. If a substimulus sampled from the rich schedule had been presented on the center key, the first left-key peck was reinforced (a hit or left correct); a right was not reinforced (a miss or right error). If a substimulus 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).

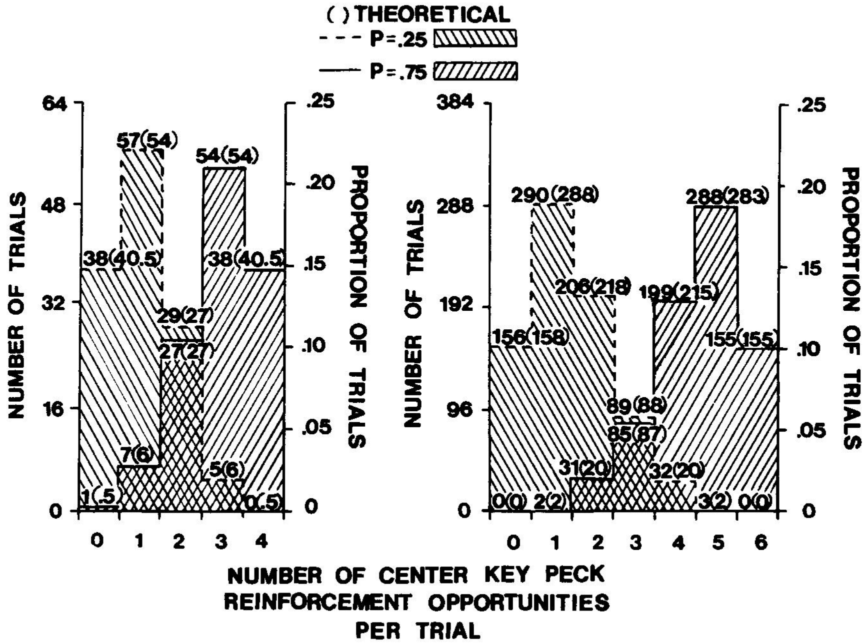
Three different standard cycle lengths were used—2 seconds, 3 seconds, and 4 seconds ($T = 2, 3, 4$). Each was run until the birds stabilized. In daily sessions the standard (base) cycle length was of standard length (i.e., T multiplied by 1) on 224 trials, doubled (T multiplied by 2) on 16 probe trials, or tripled (T multiplied by 3) on another 16 probe trials. The position of the probe trials within a session was randomly distributed. The frequency of the occurrence of the possible substimuli for standard and probe trials is shown in the left panel of Figure 3-2. The modal density of the rich schedule is 3; the modal density of the lean schedule is 1.

Stimulus Description

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 "substimulus," here called S_{rich} and S_{lean}, respectively; sampling considerations are of no concern. These stimuli are equivalent to cued random interval (RI) or variable interval (VI) schedules with a T-second limited hold and average intervals equal to $4/3\,T$ seconds and $4T$ seconds.

Figure 3–2. The distribution of reinforcement for the first center-key peck ($S^{R+}|R_C$) for four- (the left panel) and six-cycle substimuli (the right panel) 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 $p(S^{R+}|R_C)$ are 0.75 and 0.25 for the four-cycle substimuli and 0.767 and 0.233 for the six-cycle substimuli.



2. At the "molar" level, all substimulus samples with the same number of reinforcements 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 zero to four reinforcers per substimulus, giving rise to five such "stimuli," designated S_{N_d}. This level of analysis distinguishes between the number of reinforcements within a sample stimulus but not between the particular T-second cycles within that $4T$-second stimulus period on which those reinforcements are programmed.

3. At the "molecular" level, the definition of a "stimulus" involves the pattern of reinforcement in a substimulus. Each substimulus is represented as a four-digit binary number. A number such as 0001 indicates a sample with three cycles without reinforcement opportunities followed by one cycle with an opportunity. The substimuli, S_n, are numbered from 0000 to 1111. The leftmost digit represents the cycle furthest in time from choice, and the rightmost represents the cycle immediately before choice.
4. At the "micro" level, the definition of a "stimulus" depends on whether or not there is a reinforcement opportunity on a particular cycle, T-seconds preceding choice, irrespective of what is programmed for its neighbors.

There are sixteen possible combinations in a 4-cycle substimulus. The probability that a particular set of four 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.

A molecular substimulus, even 0000 or 1111, could occur with either schedule stimulus. However, the greater the number of cycles having a reinforcement opportunity, the greater the likelihood that the rich schedule was in effect. Therefore the lean (or rich) schedule was more likely to be in effect when 0 or 1 (or 3 or 4) reinforcement opportunities were presented on a trial (respectively). Substimuli with 2 reinforcement opportunities occurred about equally often given either schedule. Birds were run until their performances stabilized. The data reported here were collected in the five sessions following stabilization.

Results

The results of the present study encompass several levels of description along two stimulus dimensions, one reflecting the number and distribution of reinforcers in a sample substimulus and the other, cycle length, which determines the temporal distance of reinforcers to choice. For the reinforcement dimension, at the macro level, the sensitivity of the birds to the difference between two different reinforcement distributions will be described for the nine possible combinations of cycle lengths. At the molar, molecular, and micro levels, the decision rules, which describe the relationships between choice and particular substimulus parameters such as number of reinforcements in a substimulus and their relative distance from the choice point, can be studied in such a way as to explain the birds' sensitivity to and perceived value of the substimulus.

For the second dimension, cycle length, two operations were used to generate the temporal distance of a reinforcer in a given substimulus from choice. Each can be used to examine the effects of cycle length on perceived value. One operation established one of three standard or base cycle lengths over a large number of sessions (stable change). The second operation either doubled or tripled the standard cycle length on probe trials (momentary change). Presentation of results will proceed from most aggregated to least aggregated, along both dimensions. Sensitivity will be discussed before perceived density, although the former may depend on the latter.

First, the effects of lengthening cycles on d' and $p(L)$ are examined at the macro level. The leftmost panels in Figure 3-3 show the p (Hit) and p (False Alarm) coordinates for each value of overall cycle length, irrespective of how the cycle length was obtained (i.e., from a standard or a probe trial). These are superimposed on receiver-operating-characteristic (ROC) curves, which were calculated for a number of levels of sensitivity, according to the continuous binomial method detailed in Commons (1979). There are two values associated with each point. These are d', which reflects sensitivity to the difference between the density of the rich and lean stimuli, and $p(L)$, the probability of a left-key peck, which is the tendency to indicate that the substimulus came from the rich schedule. The value, $p(L)$, is a measure of perceived reinforcement density (bias). From an examination of the position of the points relative to the ROC curves for sensitivity and to the negative diagonal for bias, one can see the effects of increasing cycle length or of the number of cycles on sensitivity and perceived value.

Both d' and $p(L)$ decreased as cycle length increased. The decrease in d' was indicated by the points falling nearer the positive diagonal, the isosensitivity curve that reflects no sensitivity. The decrease in perceived density was indicated by points falling closer to the x axis and farther from the negative diagonal. On probe trials, both sensitivity and perceived density should decrease together, because assigning a lower density to a sample than it has decreases the hit rate (see Nevin, Chapter 1, for a definition) without decreasing the false alarm rate.

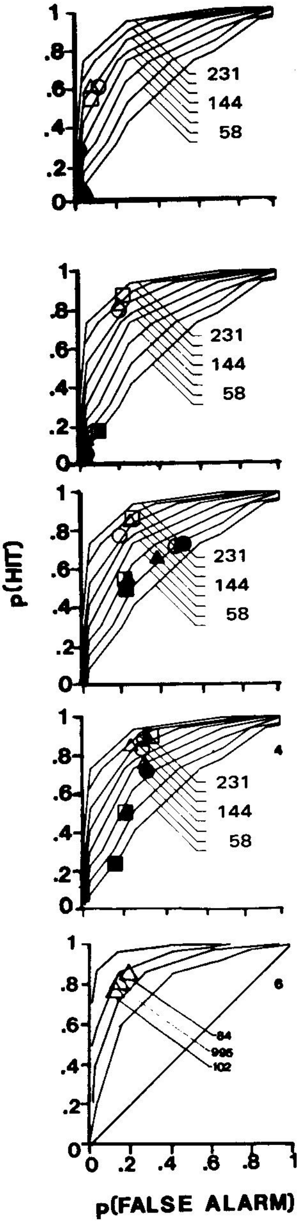
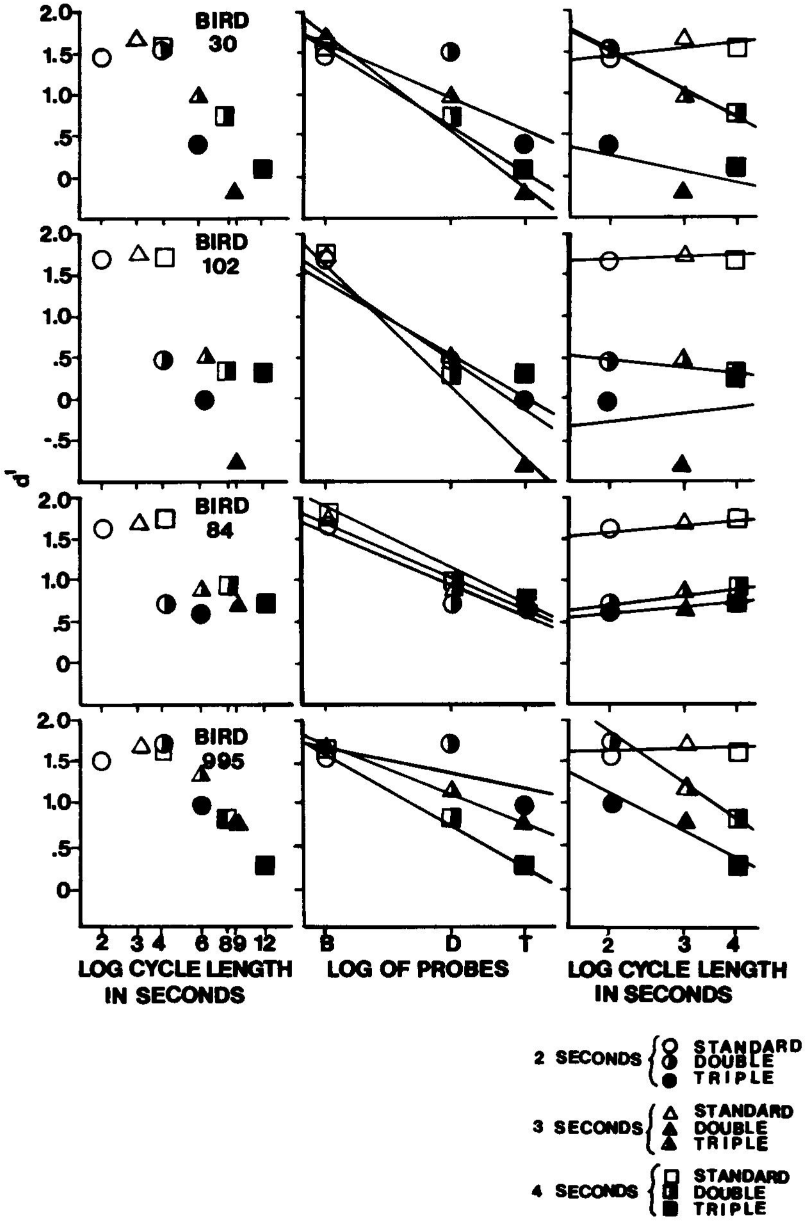


Figure 3–3. Column 1.: Isosensitivity curves for four birds (top four panels) discriminating four-cycle samples and three birds discriminating six-cycle samples (bottom panels) are shown. The probability of a hit, *p*(Hit), is equal to the number of correct left-key pecks divided by the number of possible correct left-key pecks. The probability of a false alarm, *p*(False Alarm), is equal to the number of incorrect left-key pecks divided by the number of possible correct right-key pecks. The solid lines show isosensitivity curves for eight values of d', a measure of sensitivity. The top curve would be obtained if the subject followed an ideal decision rule derived from the continuous binomial distributions in Figure 3-2. The closer points fall to the 0,0 corner the smaller the perceived density of the substimulus. Throughout the rest of the figures a common symbol code is used. Circles, triangles, and squares represent 2-, 3-, and 4-second base (standard) cycle length. The open, half-filled, and completely filled symbols represent trials with standard cycle length (open) and probe trials obtained by doubling (half-filled) and tripling (completely filled) standard cycle lengths. The second, third, and fourth columns plot d', a sensitivity measure found from the inverse to the normal probability distribution as an approximation to the binominal. Column 2 shows d' versus cycle length irrespective of how obtained; whereas Column 3 shows d' as a function of the base, doubling of base, and tripling of base cycle length on probe trials; and Column 4, as a function of base cycle length. Best fit lines indicate the other parameter.

Figure 3-6. continued



The general decrease in *d'* as a function of increasing cycle length is more clearly seen in the second column of panels in Figure 3-3. The *d'* values were found by assuming the normal approximation to the binomial rather than the continuous binomial model reported in the leftmost panel of Figure 3-3. Here, the sensitivity, *d'*, is plotted against log cycle length, the time dimension being logged to make the relation more linear. Bias is examined separately in Figure 3-4. For Birds 30, 102, and 995, *d'* was slightly higher at the 3-second cycle length than at either the 2- or 4-second base cycle lengths. From our observations, the birds had trouble pecking the center key and picking up the programmed reinforcer and then repositioning themselves in front of the center key all within two seconds. The cycle lengths greater than 4 seconds were obtained by either doubling or tripling different standard (base) cycle lengths. When disregarding how a given cycle length was obtained, this graph clearly shows that as cycle length increased beyond 4 seconds, there were large decreases in *d'* for Birds 30 and 995, a less uniform but large drop for 102, and no consistent pattern for Bird 84.

The panels in the third and fourth columns of Figure 3-3 segregate the information from the second column so that the effects of the two ways of producing different cycle lengths may be clearly seen. Sensitivity to the differences between the two schedules did not decrease in a simple fashion because the two mechanisms for lengthening cycles did not produce equal effects. Momentary increases in cycle length on probe trials produced the largest decrease in sensitivity, as shown by the negative slopes in the panels of the third column and the different heights of lines in the panels of the fourth column. Generally, the longer standard cycle lengths, when doubled and tripled, produce lower sensitivities, especially for Birds 30 and 995, indicating an interaction between probe ratio and standard cycle length. This can be seen by the generally descending order of sensitivity from the logged standard (base) cycle lengths to the doubled and tripled probes in column 3 and by the difference in the slope of the lines in column 4. One critical pair of points in column 4 occurs for the 6-second cycle length, because it can be obtained by either doubling the 3-second standard or tripling the 2-second standard. The former point is a half-filled triangle, and the latter is a fully shaded circle. The doubling of 3 seconds produces less of a sensitivity loss than the tripling of 2 seconds.