JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1989, 51, 87-99 NUMBER 1 (JANUARY)

THEORIES OF PROBABILISTIC REINFORCEMENT

JAMES E. MAZUR

SOUTHERN CONNECTICUT STATE UNIVERSITY

In three experiments, pigeons chose between two alternatives that differed in the probability of reinforcement and the delay to reinforcement. A peck at a red key led to a delay of 5 s and then a possible reinforcer. A peck at a green key led to an adjusting delay and then a certain reinforcer. This delay was adjusted over trials so as to estimate an indifference point, or a duration at which the two alternatives were chosen about equally often. In Experiments 1 and 2, the intertrial interval was varied across conditions, and these variations had no systematic effects on choice. In Experiment 3, the stimuli that followed a choice of the red key differed across conditions. In some conditions, a red houselight was presented for 5 s after each choice of the red key. In other conditions, the red houselight was present on reinforced trials but not on nonreinforced trials. Subjects exhibited greater preference for the red key in the latter case. The results were used to evaluate four different theories of probabilistic reinforcement. The results were most consistent with the view that the value or effectiveness of a probabilistic reinforcer is determined by the total time per reinforcer spent in the presence of stimuli associated with the probabilistic alternative. According to this view, probabilistic reinforcers are analogous to reinforcers that are delivered after variable delays.

Key words: reinforcer probability, reinforcer delay, intertrial interval, adjusting schedule, conditioned reinforcement, key peck, pigeons

---

In research on choice, there has been considerable interest in how subjects, both human and nonhuman, respond in situations in which the consequences of their actions are uncertain. With humans, the research frequently involves presenting subjects with written descriptions of hypothetical choices involving different amounts of money (e.g., $100 with a probability of 1.0 versus $200 with a probability of .5) and asking subjects to choose as if their decisions were real (e.g., Kahneman & Tversky, 1979; Machina, 1987). With animals, the typical procedure is to give the subjects repeated exposure to pairs of alternatives (e.g., one food pellet with a probability of 1.0 versus two pellets with a probability of .5) and to record their choices after they have been exposed to sufficient trials to learn about the two alternatives (e.g., Logan, 1965; Young, 1981). Many training trials are necessary in this procedure, and this is not simply because animals are slower to learn than people. Any subject, no matter how intelligent, would need multiple trials before being able to estimate the probability of an uncertain event with reasonable accuracy.

In both types of research, it is frequently assumed that probability of reinforcement is a distinct dimension to which subjects can respond, in much the same way that they can respond to other characteristics of a reinforcer such as its magnitude, quality, or delay. For instance, expected utility models assume that a human subject multiplies the *subjective utility* of each possible outcome by its probability to obtain the expected utility of a particular alternative, and then chooses whichever alternative has the higher expected utility (see Lee, 1971). For animal subjects, theorists do not suggest that the subjects actually perform calculations; nevertheless, it is frequently assumed that probability and amount of reinforcement combine multiplicatively to influence the subject's choice (e.g., Battalio, Kagel, & McDonald, 1985; Caraco, Martindale, & Whittam, 1980).

Rachlin, Logue, Gibbon, and Frankel (1986) proposed a different approach to probabilistic reinforcers, one that was intended to apply to both animals and people. In essence, their theory states that probabilistic reinforcers are functionally the same as delayed reinforcers,

---

This research was supported by Grant MH 38357 from the National Institute of Mental Health. I thank David Coe and Christina Berry for running the animals in these experiments, and David Cook for his help with the data analyses. Correspondence concerning this article should be sent to James E. Mazur, Psychology Department, Southern Connecticut State University, New Haven, Connecticut 06515.

and that they should be reinterpreted as such. That is, when a reinforcer is delivered with a probability less than 1.0, the effect on the subject's behavior is the same as if the reinforcer were delayed. Rachlin et al. offered a specific equation for translating probabilities into equivalent delays:

$$D = \frac{c + t}{p} - t, \qquad (1)$$

where D is the average delay to a reinforcer delivered with a probability of p. That is, D is the average time a subject would have to wait for each reinforcer if the probabilistic alternative were chosen each time. The variable t is the duration of any intertrial interval (ITI), and c is the time from the start of a trial until either the reinforcer is delivered or, on negative trials, until the trial ends without reinforcement and the ITI begins. Thus c includes the time it takes the subject to make a choice, the time needed to fulfill any additional response requirements, and the duration of any other delays between the choice response and reinforcement (or trial termination on nonreinforced trials). For example, suppose that an animal responds for food reinforcers in a discrete-trials procedure in which p is .2, the time it takes the animal to respond on each trial (c) is about 3 s, and t is 10 s. Because on average it will take five trials to obtain a reinforcer, Equation 1 states that the expected delay to reinforcement is 55 s (five trials of 3 s each, plus four ITIs of 10 s each).

The calculation of average time to reinforcement in Equation 1 is incontrovertible, but what makes the analysis of Rachlin et al. (1986) a testable theory is its assertion that the probabilistic reinforcer in the above example will be the behavioral equivalent of a reinforcer delayed 55 s. Because of this assumption about the relation between reinforcer probability and delay, D will be called the *effective delay* for a reinforcer that occurs with a probability of p. Equation 1 makes a number of predictions that differ from those of traditional expected utility theories. One is that if a subject must choose between a small but certain reinforcer and a larger probabilistic reinforcer, preference for the probabilistic reinforcer will decrease as the ITI grows longer. This is because D will increase as t increases for the probabilistic reinforcer, and it is well known that preference decreases as delay of reinforcement increases (e.g., Ainslie, 1974; Chung, 1965). For the certain reinforcer, D will not change as t increases, because when every trial delivers a reinforcer, the expected delay to reinforcement is simply the time it takes to complete the trial, c. In contrast, expected utility theories (e.g., Kahneman & Tversky, 1979) predict no change in preference with changes in the ITI (at least not without additional assumptions), because a change in the ITI affects neither the probability nor the amount of reinforcement for either alternative. Rachlin et al. provided some evidence supporting Equation 1 with a gambling experiment in which people chose between a large low-probability payoff and a small but nearly certain payoff.

One purpose of the present set of experiments was to test the predictions of Equation 1 with nonhuman subjects. Pigeons chose between alternatives that differed, not in amount, but in the probability of reinforcement and in the delay between choice and reinforcement. An adjusting-delay procedure (Mazur, 1984, 1987) was used to estimate indifference points—pairs of alternatives a subject selected about equally often in a choice situation. In this procedure, the delay for one alternative, the *standard alternative,* is constant throughout a condition. The delay for the other alternative, the *adjusting alternative,* changes many times a session so that the indifference point can be estimated. In Experiments 1 and 2, pigeons chose between a certain reinforcer and a probabilistic reinforcer, and the duration of the ITI was varied across conditions. These experiments examined the prediction of Equation 1 that preference for the probabilistic reinforcer should decrease as ITI increases. The procedures in these experiments were very similar to those of a previous study (Mazur, 1985), but ITI was not varied in that study.

In Experiment 3, the stimuli present during the delay that followed each choice of the probabilistic reinforcer were varied. Specifically, in some conditions the same stimulus was present on those trials in which a reinforcer would be delivered and on those in which none would be delivered. In other conditions, different stimuli were present on reinforced and nonreinforced trials. As will be shown, the details of such stimulus arrangements can have sub-

Table 1

Order of conditions and mean adjusting delays at the indifference point in Experiment 1. All durations are in seconds.

Subject	Order 1 and 2	Order 3 and 4	p_s	Trial duration	Adjusting delay 1	2	3	4	Mean
1	1	1.0	60	3.21	6.46	5.07	8.57	5.83
2	2	.2	60	8.32	22.82	21.13	26.27	19.64
3	3	.5	60	6.17	9.44	6.79	21.59	11.00
4	4	.2	60	7.69	32.71	11.21	28.93	20.14
5	6	.5	30	6.06	6.80	8.09	11.86	8.20
6	5	.5	90	7.19	9.80	10.25	17.63	11.22
7	8	.5	30	6.84	8.77	14.46	7.44	9.38
8	7	.5	90	5.50	7.23	8.34	14.19	8.82

stantial effects of choice, even when the delays, probabilities, and ITIs for the two alternatives remain the same.

EXPERIMENT 1

In most of the conditions of this experiment, the standard alternative delivered a reinforcer with a probability of .5 after a 5-s delay. The adjusting alternative delivered a reinforcer on every trial after an adjustable delay. The purpose of this experiment was to vary t across conditions and determine whether the indifference point changed as a result. However, if t were the same for both the standard and adjusting alternatives, the rate of trial presentations would differ for the two alternatives whenever their delays were different. To minimize variation in the rate of trial presentations within a condition, the time from a choice response until the start of the next trial (which will be called *trial duration*) was kept constant within each condition. In different conditions, the trial duration was 30 s, 60 s, or 90 s.

Because the trial duration, the two probabilities, and the standard delay were constant in this procedure, it is a simple matter to derive predictions from Equation 1. By definition, the values of D_s and D_a will be equal at the indifference point (where the subscripts refer to the standard and adjusting alternatives, respectively). Let us ignore the small amounts of time the subjects need to make their choice, because in practice these durations are so small that they would not affect the predictions appreciably. With a trial duration of 30 s for the standard alternative, Equation 1 predicts that indifference would be reached with an adjusting delay of 35 s, if such a long delay were possible. However, in this condition, the maximum adjusting delay was limited to 28 s to keep the trial duration for the adjusting alternative equal to 30 s as well. Equation 1 therefore predicts that the adjusting delay will reach its longest possible duration. Similarly, for conditions with trial durations of 60 s and 90 s, Equation 1 predicts that the adjusting delay will reach 58 s and 88 s, respectively—the maximum possible durations in these two conditions.

METHOD

Subjects

Four White Carneau pigeons were maintained at about 80% of their free-feeding weights. All had previous experience with a variety of experimental procedures.

Apparatus

The experimental chamber was 30 cm long, 30 cm wide, and 33 cm high. Three response keys, each 2.5 cm in diameter, were mounted in the front wall of the chamber, 20.5 cm above the floor. A force of approximately 0.1 N was required to operate each key, and each effective response produced a feedback click. A hopper below the center key provided controlled access to mixed grain, and when grain was available, the hopper was illuminated with two 6-W white lights. Six 6-W lights (two white, two red, and two green) were mounted above the wire-mesh ceiling of the chamber. The chamber was enclosed in a sound-attenuating chamber that contained an air blower for ventilation

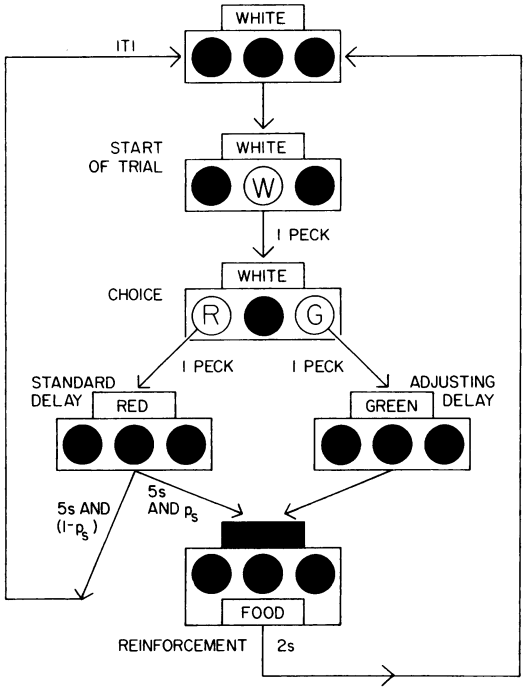

Fig. 1. A schematic diagram of the different sequences of events that could occur on a choice trial in Experiment 1.

and a speaker producing continuous white noise to mask extraneous sounds. A PDP-8® computer in another room was programmed in SuperSKED® to control the stimuli and record responses.

Procedure

The experiment consisted of eight conditions that differed in trial duration and in p_s, the probability of reinforcement for the standard alternative. In the first four conditions, trial duration was kept constant and p_s was varied to assess the effects of different probabilities. In the last four conditions, p_s was constant at .5, and trial duration was varied. Table 1 lists the trial durations and values of p_s for each condition. Notice that in the last four conditions, the order in which the conditions were studied was different for Subjects 3 and 4.

Each session lasted for 64 trials or for a maximum duration that depended on the trial duration. For trial durations of 30 s, 60 s, and 90 s, the maximum session durations were, respectively, 40 min, 80 min, and 120 min. Each block of four consecutive trials consisted of two forced trials followed by two choice trials. Figure 1 illustrates the possible sequences of events that could occur on a choice trial. After an ITI during which only the white houselights were illuminated, the center key was illuminated with white light to start a trial. A single peck on the center key was required to begin the choice period. The purpose of this center peck was to make it more likely that the subject's head was equidistant from the two side keys when the choice period began. A peck on the center key darkened this key and illuminated the two side keys, one green and one red. The locations of these colors (left key or right key) varied randomly from trial to trial to control for position preferences.

The red key was always the standard key and the green key was the adjusting key. A peck on the red key extinguished both side keys and initiated a 5-s delay, during which the red houselights were lit instead of the white houselights. At the end of the standard delay, the red houselights were extinguished and either (a) with a probability of p_s, grain was presented for 2 s, followed by the next ITI, or (b) with a probability of $1 - p_s$, there was no food presentation, and the next ITI began immediately. A peck at the green key during the choice period extinguished both side keys and led to the adjusting delay, during which the green houselights were lit. The adjusting delay was always followed by 2-s access to grain and then the ITI. On all trials, the duration of the ITI was varied so that the trial duration was constant.

The procedure on forced trials was the same as on choice trials, except that only one side key was lit, red or green, and a peck on this key led to the appropriate delay. A peck on the opposite key, which was dark, had no effect. Of every two forced trials, one involved the red key and the other the green key. The temporal order of these two types of trials varied randomly.

After every two choice trials, the delay for the adjusting key might be changed. If a subject chose the adjusting key on both choice trials, the adjusting delay was increased by 1 s. If the subject chose the standard key on both trials, the adjusting delay was decreased by 1 s. If the subject chose each key on one trial, no change was made in the adjusting delay. In all three cases, this adjusting delay remained in effect for the next block of four trials. At the start

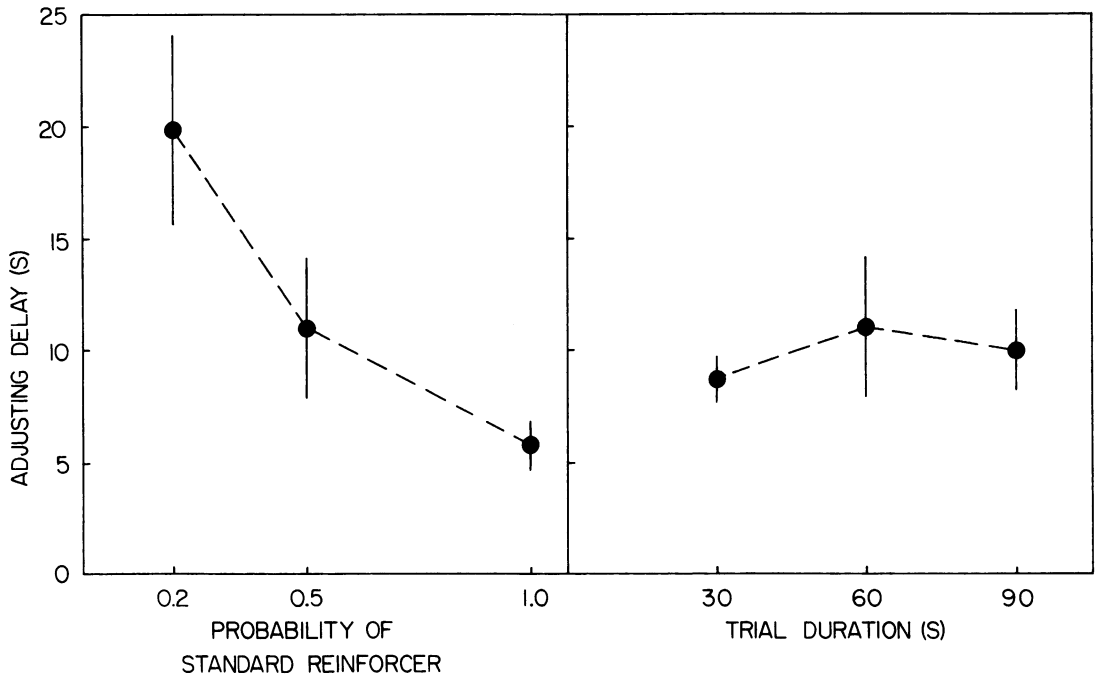

Fig. 2. Mean indifference points (and standard errors of the mean) from the different conditions of Experiment 1. Results from conditions with a trial duration of 60 s are plotted as a function of the standard reinforcer probability in the left panel. Results from conditions with a standard probability of .5 are plotted as a function of trial duration in the right panel. The middle points in the two panels are based on the same data.

of each session, the adjusting delay was determined by the above rules as if it were a continuation of the preceding session.

The first two conditions lasted for a minimum of 20 sessions, and the other conditions lasted for a minimum of 12 sessions. After the minimum number of sessions, a condition was terminated for each subject individually when several stability criteria were met. To assess stability, each session was divided into two 32-trial blocks, and for each block the mean delay on the adjusting key was calculated. The results from the first two sessions of a condition were not used, and a condition was terminated when the following three criteria were met, using the data from all subsequent sessions: (a) Neither the highest nor the lowest single-block mean of a condition could occur in the last six blocks of the condition. (b) The mean adjusting delay across the last six blocks could not be the highest or the lowest six-block mean of the condition. (c) The mean delay of the last six blocks could not differ from the mean of the preceding six blocks by more than 10% or by more than 1 s (whichever was larger).

Results and Discussion

In every condition, the mean adjusting delay from the six half-session blocks that satisfied the stability criteria was used as an estimate of the indifference point. Table 1 shows these durations for each subject and for the group. The group results are also presented in Figure 2, where the results are averaged across replications for those conditions that were repeated. The left panel shows that, with trial duration unchanged (Conditions 1 through 4), the mean adjusting delay decreased systematically as the probability of reinforcement for the standard alternative increased. As can be seen in Table 1, this decreasing pattern was found for every subject, although the range of adjusting delays varied considerably across animals. These results demonstrate that the subjects were sensitive to the changes in the standard probability of reinforcement.

The right panel of Figure 2 shows the results from Conditions 5 through 8 (and again from Condition 3), where the standard probability was .5 and trial duration varied. Two

aspects of the data are noteworthy. First, the indifference points were much shorter than those predicted by Equation 1. Recall that, according to Equation 1, the indifference points should range from 28 s to 88 s, whereas the actual indifference points averaged about 10 s. Second, no trend toward longer adjusting delays with longer trial durations was observed. The slightly larger group means for trial durations of 60 s and 90 s were entirely attributable to one animal, Subject 4. The results from the other 3 subjects showed no increases at all (see Table 1).

These results clearly contradict Equation 1. However, this does not necessarily mean that the general concept proposed by Rachlin et al. (1986)—that probabilistic reinforcers are the behavioral equivalent of delayed reinforcers—is incorrect. Perhaps the animals in this study responded to the probabilistic reinforcers as they would to delayed reinforcers, but they did not treat ITI time as part of the delay. Within the framework of the Rachlin et al. theory, ignoring ITI time in this experiment could be justified on the grounds that the ITIs were associated with different discriminative stimuli (the white houselights) and they were therefore easily discriminable from the delay periods. Suppose D_s, the effective delay for the standard alternative, is defined as the average time, cumulated across trials, that the subject spent in the presence of the stimuli associated with the standard delay (the red keylight and red houselight) before a reinforcer was delivered. Similarly, suppose D_a is the average time spent in the presence of the adjusting delay stimuli (the green keylight and green houselight) before a reinforcer was delivered. The time spent in the presence of the white houselights (which were on during the ITI after both standard and adjusting trials) is simply not included for either alternative. Under these assumptions, Equation 1 reduces to

$$D = \frac{c}{p}. \qquad (2)$$

D is now not the total time before reinforcement but the average time spent in the presence of the delay stimuli before reinforcement. As can be seen, Equation 2 predicts an inverse relation between reinforcer probability and reinforcer delay.

Equation 2 makes two predictions that are consistent with the results of this experiment. First, it predicts that ITI duration should have no effect on choice. Second, the predicted indifference point for the conditions with p_s = .5 was fairly close to the mean result. If c includes the durations of both the red keylight and the red houselight, it is necessary to know the average response latencies on the standard and adjusting keys in order to make a precise prediction of the indifference point. In conditions with p_s = .5, the mean response latencies on the standard and adjusting keys were 1.0 s and 1.1 s, respectively. With these values, Equation 2 predicts an indifference point of 10.9 s, whereas the mean indifference point across the five conditions with p_s = .5 was 9.7 s.

For the conditions with p_s = .2, the predictions of Equation 2 were less accurate. Response latencies were 1.9 s and 1.5 s on the standard and adjusting keys, respectively. With these latencies, Equation 2 predicts an indifference point of 33.0 s, whereas the average obtained indifference point was 19.9 s. There was, however, a great deal of variability across subjects and sometimes across the two conditions with p_s = .2.

Although the predictions of Equation 2 were not entirely accurate at a quantitative level, they were certainly more accurate than those of Equation 1. Nevertheless, it is still possible that the latter equation is partly correct. That is, perhaps ITI has some small effect on choice when probabilistic reinforcers are involved, but not to the extent predicted by Equation 1. The effect might have been too small to be observed in this experiment. If so, an effect of ITI duration might be observed with probabilities smaller than .5, because in Equation 1, D_s increases as probability decreases. To provide a potentially more sensitive test of the effects of ITI duration, Experiment 2 measured indifference points when p_s was .2 with two different trial durations, 40 s and 90 s.

EXPERIMENT 2

METHOD

Subjects and Apparatus

The subjects were 4 different White Carneau pigeons maintained at approximately 80% of their free-feeding weights. All had previous experience with a variety of experimental pro-

cedures. A different experimental chamber, with slightly different dimensions but otherwise similar to that of Experiment 1, was used.

Procedure

As in Experiment 1, the adjusting alternative delivered a reinforcer on every trial. The standard delay was always 5 s, and the standard probability of reinforcement was .2. In two conditions, the trial duration was 40 s, and in another two conditions the trial duration was 90 s. The order of these two types of conditions was counterbalanced across subjects. The same stability criteria were used to terminate conditions as in Experiment 1, with Condition 1 lasting for a minimum of 20 sessions and the other conditions a minimum of 12 sessions. In all other respects the procedures were the same as in Experiment 1.

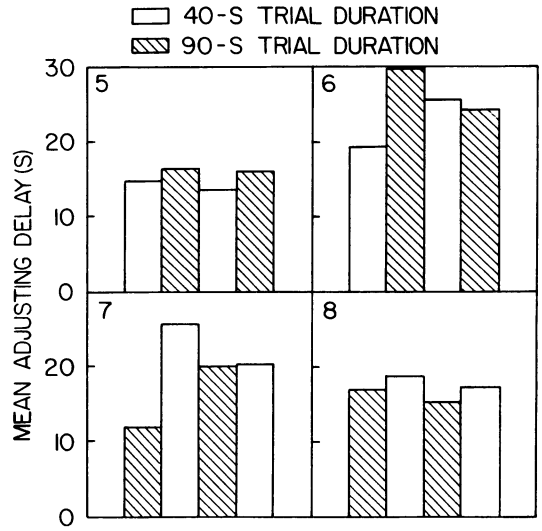

Fig. 3. Mean adjusting delays in the four conditions of Experiment 2 are shown for each subject. The results are shown in the order in which the conditions were presented to each subject.

Results and Discussion

Figure 3 presents the indifference points obtained in each condition for each subject and shows that the indifference points did not differ systematically as a function of trial duration. The mean adjusting delay was 19.3 s with the 40-s trial duration and 19.1 s with the 90-s trial duration.

These results are similar to those of Experiment 1 in two ways. First, contrary to the prediction of Equation 1, trial duration had no effect on choice. Second, the actual indifference points were shorter than predicted by Equation 2. With standard and adjusting response latencies averaging 1.4 s and 2.4 s, respectively, Equation 2 predicts an indifference point of 29.6 s, compared to the mean obtained indifference point of 19.1 s.

The indifference points obtained in the similar study of Mazur (1985) were also shorter than those predicted by Equation 2. With $p_s = .5$, indifference points averaged 7.8 s and were less than 11.0 s, the duration predicted by Equation 2, in eight of eight cases. With $p_s = .2$, indifference points averaged 12.0 s and were less than the predicted duration of 29.0 s in eight of eight cases. Combined with the results of Experiments 1 and 2, these findings clearly show that the indifference points predicted by Equation 2 are too large.

Although Equations 1 and 2 have difficulty accounting for the results of these studies, the basic assumption of Rachlin et al. (1986)—that probabilistic reinforcers are the equivalent of delayed reinforcers—can be applied in another way that appears to be more consistent with the data. This analysis hinges on the observation that the actual amounts of time spent in the presence of the standard stimuli varied substantially from one standard reinforcer to the next. For instance, with $p_s = .2$, the average time spent in the presence of the standard stimuli might be about 30 s (assuming response latencies of about 1 s). However, because the number of standard trials needed to obtain a reinforcer could range from one to more than 10, the actual times spent in the presence of standard stimuli could be as little as 6 s or more than 60 s. Because of this variability, if probabilistic reinforcers are to be interpreted as delayed reinforcers, it seems reasonable to treat them as reinforcers with variable rather than fixed delays (cf. Hamm & Shettleworth, 1987). It is known that animals prefer schedules with variable requirements over those with fixed requirements (e.g., Fantino, 1967; Herrnstein, 1964), and more specifically, that they prefer variable delays to reinforcement over fixed delays (e.g., Cicerone, 1976; Mazur, 1984; Rider, 1983). Perhaps the problem with Equation 2 is that it implicitly treats a probabilistic rein-

forcer as one delivered after a fixed delay (that is, the calculation of D_s does not take trial-to-trial variability into account). This might explain why the obtained indifference points were consistently shorter than predicted by Equation 2: These shorter indifference points indicate a greater than predicted preference for the probabilistic reinforcer; this preference might occur because of the variable number of standard trials needed to obtain each standard reinforcer.

In order to make this hypothesis more precise, one must know more about how animals choose between fixed and variable delays. Mazur (1984, 1986) has found that the following equation provides reasonably accurate predictions for choice between fixed and variable delays:

$$V = \sum_{i=1}^{n} P_i \frac{1}{1 + KD_i}. \qquad (3)$$

V is the value of a schedule composed of n different possible delays to reinforcement, and P_i is the probability that a delay of D_i seconds will occur on any given trial. K is a free parameter that determines how rapidly V declines with increasing values of D_i. (Note the distinction between P_i, the probability that a certain delay will occur, and p_i, the probability of reinforcement on a given trial. In Mazur's 1984 and 1986 studies, p_i was always 1.0.) Mazur was able to make predictions for choices between any two schedules of delayed reinforcers by assuming that V would be the same for the two schedules at the indifference point. Although K was treated as a free parameter in some of this work, good predictions were usually obtained by setting K equal to 1, and this value of K will be used in the following calculations.

Equation 3 was used to make predictions for the present experiments by defining D as the time spent in the presence of the red (or green) stimuli before each reinforcer. Unlike Equation 2, however, Equation 3 used a distribution of values for D rather than one average value. The calculations based on Equation 3 can be illustrated by considering the conditions in Experiment 1 in which p_s was equal to .5. Values of P_i were obtained using the same half-session blocks used to estimate the indifference points. Across all subjects and conditions with $p_s = .5$, a total of 902 standard reinforcers were delivered in these half-sessions. Of these, 451 were delivered on the first choice of the standard alternative following a standard reinforcer. Therefore P_1, the probability of receiving a reinforcer on the first choice of the standard key since the previous standard reinforcer, was .5 (451/902). D_1, the time spent in the presence of the standard stimuli on these occasions, was estimated to be 6.0 s (which includes the response latency plus the 5-s standard delay). Other values of P_i and D_i were calculated in an analogous manner. For instance, because 143 of the standard reinforcers were delivered on the third choice of the standard key following the previous standard reinforcer, P_3 was .16, and D_3 was estimated to be 18.0 s (three times D_1). By using these values of P_i and D_i in Equation 3, it was estimated that V_s, the value of the standard alternative, was 0.102. Assuming that $V_s = V_a$ at the indifference point, Equation 3 predicts an indifference point of 7.7 s for these conditions. (This analysis ignores the fact that the adjusting delay changed gradually over trials, which simplifies the calculations but does not alter the predictions significantly.)

Similar calculations were made for the conditions with $p_s = .2$ in Experiments 1 and 2, and for the different conditions of Mazur's (1985) study. This earlier experiment used six different combinations of probabilities for the standard and adjusting alternatives, including four in which both p_s and p_a were less than 1.0. Figure 4 plots the mean indifference points from the three experiments as a function of the predictions of both Equation 2 and Equation 3. With K set equal to 1 in Equation 3, both sets of predictions are parameter-free. Equation 3 accounted for 60.7% of the variance in these data points, whereas Equation 2 accounted for none (i.e., the variance from the predictions of Equation 2 was greater than the variance around the mean). All of the indifference points where shorter than predicted by Equation 2.

Figure 4 shows that the actual indifference points roughly approximated the predictions of Equation 3, but several limitations of these predictions are apparent. First, the deviations from the predictions were systematic, with the results from Mazur (1985) always shorter than predicted and those from Experiments 1 and 2 longer than predicted. Second, these parameter-free predictions cannot account for the individual differences observed in all three experiments. Within the general approach of

Equation 3, there are at least two possible explanations of these differences across subjects and experiments. First, subjects might vary in how strongly they are influenced by reinforcer delays, so that different values of K might be appropriate for different subjects. Second, subjects might have different biases that favor either the standard or adjusting alternative. These could be due to a color bias or to the fact that the adjusting delay varied whereas the standard delay did not. For example, the short indifference points of Subject 1 in Experiment 1 could be the result of a bias for the standard alternative. Notice that in Condition 1, where both alternatives delivered a reinforcer on every trial, the subject's indifference point was 3.21 s, whereas a value of 5 s would be expected in the absence of bias. Because the present set of experiments was not designed either to investigate or to account for such individual differences, they will not be discussed further.

In summary, the overall pattern of results in these experiments is inconsistent with Equation 1 (because trial duration had no effect) and systematically deviates from the predictions of Equation 2. To a first approximation, however, the results are consistent with Equation 3. If Equation 3 (or Equation 2, for that matter) is correct, and if D is interpreted as the time spent in the presence of the delay stimuli, it should be possible to alter the indifference point by changing the way these stimuli are presented. One such change was examined in Experiment 3.

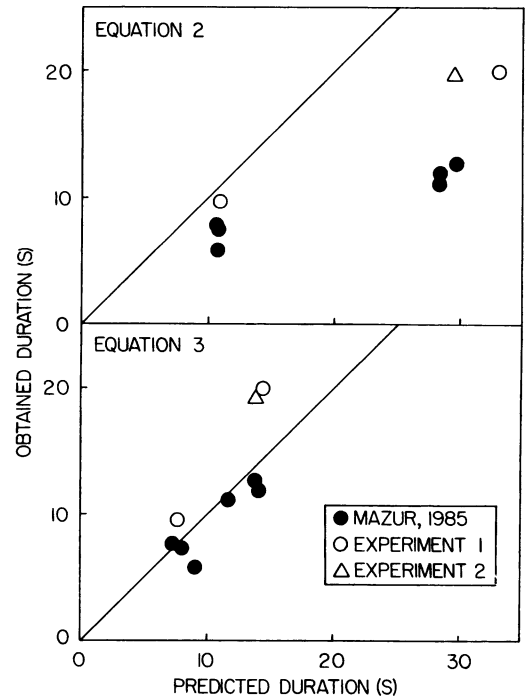

Fig. 4. For the group means from Experiments 1 and 2 and from Mazur (1985), obtained indifference points are plotted as a function of the predictions of Equations 2 and 3.

EXPERIMENT 3

In the previous two experiments, a peck at the red standard key was always followed by a 5-s presentation of the red houselight, which was followed by a reinforcer on some trials but not others. In some of the conditions of Experiment 3, a simple change was introduced: For the standard alternative, the red houselight was present on reinforced trials but not on nonreinforced trials. On nonreinforced standard trials, a peck at the red key turned the keylight off, the white houselight remained on, and the ITI began immediately. For the adjusting alternative, which delivered a reinforcer on every trial, the green houselight was always present during the delay between a choice response and reinforcement. Equations 2 and 3 both predict that preference for the standard alternative will increase under this procedure, because the time spent in the presence of the standard-key stimuli per reinforcer is decreased. An increase in preference for the standard alternative would be reflected as a decrease in the adjusting delay at the indifference point. In contrast, Equation 1 predicts no change in preference, because changing the color of the stimulus affects neither c nor t. Similarly, traditional expected utility theories predict no change in preference because p_s and p_a have not changed.

METHOD

Subjects and Apparatus

The subjects were those from Experiment 1, and the same experimental chamber was used.

Procedure

Except as described below, the procedures were the same as in Experiment 1. Throughout the experiment, the standard delay was 5 s and the trial duration was 30 s. To maintain this trial duration, the adjusting delay was not

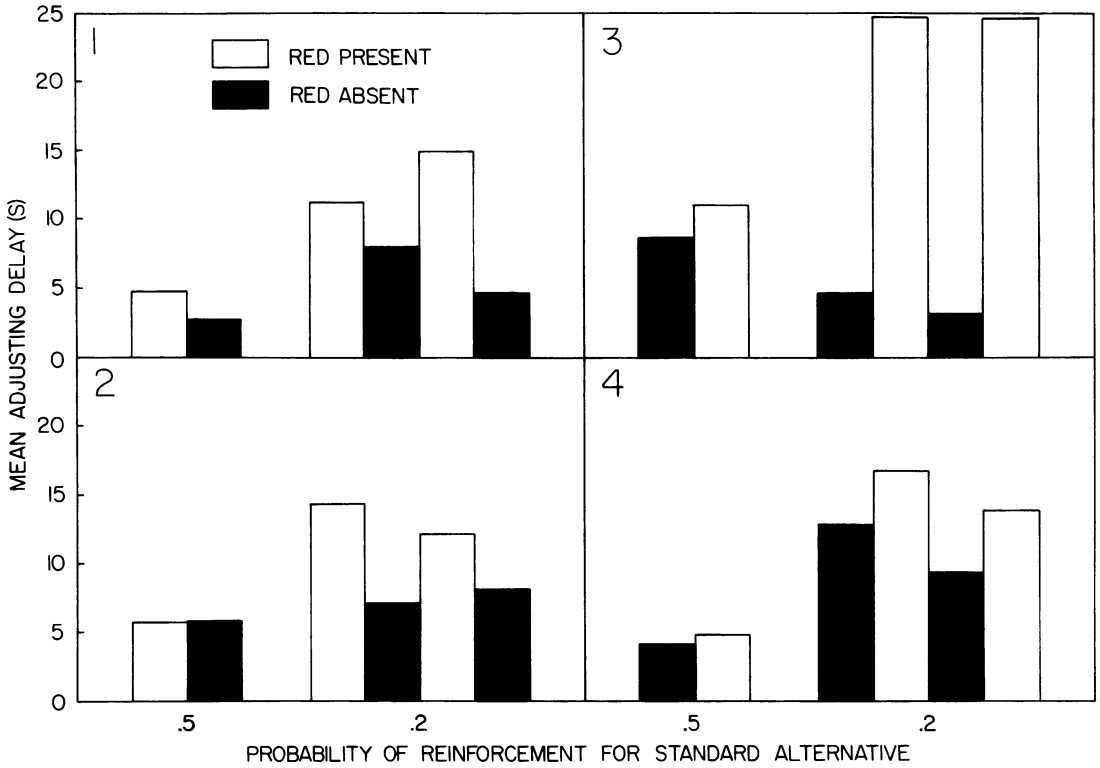

Fig. 5. Mean adjusting delays in the six conditions of Experiment 3 are shown for each subject. The results are shown in the order in which the conditions were presented to each subject.

allowed to increase above 25 s, which meant that the minimum ITI (after a 2-s reinforcer) was 3 s. The experiment consisted of three pairs of conditions. In the first pair, p_s was .5, and in the other two pairs it was .2. In one condition of each pair (with order counterbalanced across subjects), the red houselight was presented for 5 s after each choice of the standard key on both reinforced and nonreinforced trials, as in the previous experiments. Because the red houselight was present on every standard trial, this will be called the red-present condition. In the other condition of each pair (the red-absent condition), the red houselight was presented only on reinforced standard trials. On nonreinforced standard trials, a peck at the standard key extinguished the red keylight (as well as the green keylight if it had been lit), but the red houselight was not presented. Instead, the white houselight remained on, and the ITI began immediately. Because of the constant trial duration, the ITI was 23 s on reinforced standard trials and 30 s on nonreinforced standard trials. All conditions lasted for a minimum of 12 sessions except Condition 3 (in which p_s was changed from .5 to .2), which lasted for a minimum of 20 sessions. After the minimum number of sessions, the stability criteria used to terminate a condition and calculate the indifference point were the same as in Experiment 1.

Results and Discussion

Figure 5 presents the indifference points from each condition for every subject. In the first two conditions, in which p_s was .5, the presence or absence of the red houselight on nonreinforced standard trials had little effect. The indifference points were slightly shorter in the red-absent conditions for 3 of 4 subjects, and the group means were 6.6 s in the red-present condition and 5.3 s in the red-absent condition. However, houselight color had substantial effect in the conditions with p_s = .2. Figure 5 shows that for each pair of conditions and for each subject, the indifference point was

shorter in the red-absent condition than in the red-present condition. The group means were 16.6 s in the red-present condition and 7.2 s in the red-absent condition. Figure 5 also shows that this effect was much greater for some subjects than others. The indifference points for Subject 3 in the red-present condition were close to the maximum possible adjusting delay of 25 s, so these points might well have been larger had this restriction not been imposed. From a qualitative standpoint, this problem is of little importance because this subject showed the largest differences between red-present and red-absent conditions.

The shorter indifference points in the red-absent conditions indicate that, as predicted by Equation 3, subjects showed a greater preference for the standard alternative in these conditions than in red-present conditions, at least with p_s = .2. Some potential reasons for the large individual differences have already been mentioned, but because of these differences, a quantitative test of Equation 3 cannot be very informative. However, Figure 6 provides more evidence that supports Equation 3 at a qualitative level. The top panel shows the predictions of Equation 3 for each of the four types of conditions used in this experiment, and the bottom panel shows the group means. The predictions were made using the same method as for Experiments 1 and 2. Equation 3 predicts (a) little difference between red-present and red-absent conditions with p_s = .5, (b) a much larger difference with p_s = .2, and (c) not much change in the indifference points in the red-absent conditions with the two different values of p_s. As can be seen, the group results were consistent with each of these predictions.

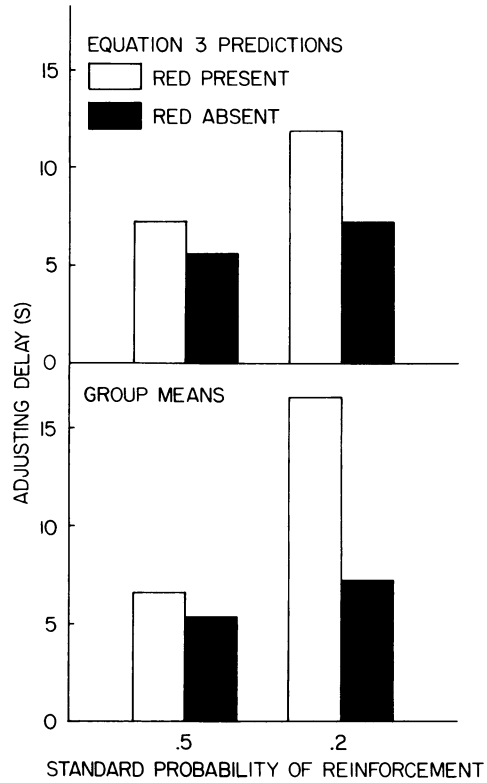

Fig. 6. The upper panel shows the predictions of Equation 3 for the four different types of conditions in Experiment 3, and the lower panel shows the group means from these four types of conditions.

GENERAL DISCUSSION

Four approaches to probabilistic reinforcement have been described in this paper—that of expected utility theory and those represented by Equations 1 through 3—and the results of the present experiments help to distinguish among them. The evidence for the expected utility approach was mixed. The failure to find an effect of trial duration in Experiments 1 and 2 was consistent with this approach, because changing the ITI does not affect the probability of either alternative. However, two of the other theories discussed also predict this lack of an effect, so the support for the expected utility approach was not unique. On the negative side, the difference between the red-present and red-absent conditions in Experiment 3 was not predicted by the expected utility approach, because the probabilities of reinforcement were the same in these two conditions.

The results were most damaging for Equation 1, which states that preference for a probabilistic reinforcer should be strongly dependent upon overall trial duration. Trial duration was varied in Experiments 1 and 2 over a range which, according to Equation 1, should have produced dramatic changes in the indifference points, but no such changes were found. Thus, although there is some evidence that the time between trials affects the behavior of human subjects choosing between reinforcers of different probabilities (Rachlin et al., 1986), the present results suggest that pigeons' choices are not affected by this variable, at least under

some conditions. This finding implies that if probabilistic reinforcers are to be viewed as the behavioral equivalent of delayed reinforcers, there are cases in which it is necessary to use some measure of delay that does not include the time between trials.

Equation 2 offers one such measure. This equation excludes ITI duration from consideration, and it asserts that the relevant variable is the *average* amount of time spent in the presence of the stimuli associated with the probabilistic alternative per reinforcer. From an empirical standpoint, the main problem with this equation is that it predicted indifference points for three experiments that were consistently too large (Figure 4). From a theoretical perspective, the problem with this equation is that it does not consider the variability in the number of trials required to obtain a reinforcer (and the resultant variability in the amount of time spent in the presence of the delay stimuli per reinforcer). Given the ample evidence that animals prefer variable over fixed delays (e.g., Cicerone, 1976; Mazur, 1984), the finding that indifference points were shorter than predicted by Equation 2 makes sense. These short indifference points are indicative of a greater preference for the probabilistic reinforcer than predicted by Equation 2, and this greater preference may have been the result of the variability in D from one reinforcer to the next.

One way to take this variability into account while retaining the general approach of Equation 2 is illustrated in Equation 3. Because Mazur (1984) found that Equation 3 provided accurate predictions for choices between fixed and variable reinforcer delays, it might also predict the effects of the variable delays to reinforcement that occur when probabilistic reinforcers are used. The only new detail involved in the present application of Equation 3 was the assumption (as in Equation 2) that D represents only the cumulative time spent in the presence of the stimuli associated with the probabilistic reinforcer (the red keylight and red houselights in these experiments). Figure 4 shows that the predictions of Equation 3 were at least roughly consistent with the results from three experiments. Although the predictions were not perfect, this equation did avoid the large and systematic errors of prediction found with Equations 1 and 2, and it therefore seems to offer a better means of predicting choice between probabilistic and certain reinforcers. Notice, however, that Equation 3 still describes a method for translating from probability to delay. The support for Equation 3 was therefore consistent with the general view of Rachlin et al. (1986) that probabilistic reinforcers are the behavioral equivalent of delayed reinforcers.

The way D is defined in Equations 2 and 3 may seem counterintuitive, because it is not obvious why time spent in the presence of one stimulus (a red houselight) should decrease preference for a probabilistic reinforcer whereas time spent in the presence of a different stimulus (a white houselight) should not. However, Experiment 3 provided stronger support for this definition of D. In conditions in which the red houselight was not present on nonreinforced trials, subjects spent less time in the presence of the probabilistic-alternative stimuli per reinforcer. As predicted by Equation 3, preference for the probabilistic alternative increased in these conditions. Furthermore, the amount of change in the indifference points, averaged across subjects, was approximately the amount predicted by Equation 3 for probabilities of both .5 and .2 (see Figure 6).

One unresolved issue is when ITI duration will and will not affect a subject's choice. With human subjects, Rachlin et al. (1986) found a greater preference for a low-probability reinforcer when ITIs were short than when they were long, as predicted by Equation 1. In contrast, ITI had no effect in the present experiments. These results do not necessarily indicate a difference in the choice behavior of people and pigeons, because in another study with pigeon subjects, Mazur, Snyderman, and Coe (1985) found that ITI size had small but consistent effects on choice. Indeed, they were able to predict their results fairly well using Equation 3, but they included ITI time in all of their calculations of D. In contrast, Equation 3 could account for the results of the present set of experiments only if ITI times were excluded from the calculations.

The difference between the present experiments and the Mazur et al. (1985) experiment could be due to the way in which discriminative stimuli were presented during the ITIs. In the Mazur et al. experiment, the same stimuli were present both during the delay that preceded a reinforcer and during the ITI that followed it (blue houselights on standard trials

and orange houselights on adjusting trials). In the present experiments, the white houselights that were present during the ITIs were (a) different from the delay stimuli and (b) the same on standard and adjusting ITIs. This arrangement of stimuli may have fostered a discrimination between the delays and the ITIs, and it might account for the absence of control by ITI duration. Further research on the effects of discriminative stimuli in these choice situations is needed.

Although Equation 3 provided the best account of the results of the present experiments, its applicability to other situations involving probabilistic reinforcers is uncertain. For instance, it remains to be seen whether it can predict preference in situations where there is no delay between a choice response and outcome (reinforcement or nonreinforcement). Nevertheless, the present results suggest that there is merit in the view that probabilistic reinforcers can be treated as delayed reinforcers, and that the variability in these delays must be taken into account.

REFERENCES

Ainslie, G. W. (1974). Impulse control in pigeons. *Journal of the Experimental Analysis of Behavior,* **21,** 485–489.

Battalio, R. C., Kagel, J. H., & McDonald, D. N. (1985). Animals' choices over uncertain outcomes: Some initial experimental results. *American Economic Review,* **75,** 597–613.

Caraco, T., Martindale, S., & Whittam, T. S. (1980). An empirical demonstration of risk-sensitive foraging preferences. *Animal Behaviour,* **28,** 820–830.

Chung, S. H. (1965). Effects of delayed reinforcement in a concurrent situation. *Journal of the Experimental Analysis of Behavior,* **8,** 439–444.

Cicerone, R. A. (1976). Preference for mixed *versus* constant delay of reinforcement. *Journal of the Experimental Analysis of Behavior,* **25,** 257–261.

Fantino, E. (1967). Preference for mixed- versus fixed-ratio schedules. *Journal of the Experimental Analysis of Behavior,* **10,** 35–43.

Hamm, S. L., & Shettleworth, S. J. (1987). Risk aversion in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes,* **13,** 376–383.

Herrnstein, R. J. (1964). Aperiodicity as a factor in choice. *Journal of the Experimental Analysis of Behavior,* **7,** 179–182.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica,* **47,** 263–291.

Lee, W. (1971). *Decision theory and human behavior.* New York: Wiley.

Logan, F. A. (1965). Decision making by rats: Uncertain outcome choices. *Journal of Comparative and Physiological Psychology,* **59,** 246–251.

Machina, M. J. (1987). Decision-making in the presence of risk. *Science,* **236,** 537–543.

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

Mazur, J. E. (1985). Probability and delay of reinforcement as factors in discrete-trial choice. *Journal of the Experimental Analysis of Behavior,* **43,** 341–351.

Mazur, J. E. (1986). Fixed and variable ratios and delays: Further tests of an equivalence rule. *Journal of Experimental Psychology: Animal Behavior Processes,* **12,** 116–124.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 5. The effect of delay and of intervening events on reinforcement value* (pp. 55–73). Hillsdale, NJ: Erlbaum.

Mazur, J. E., Snyderman, M., & Coe, D. (1985). Influences of delay and rate of reinforcement on discrete-trial choice. *Journal of Experimental Psychology: Animal Behavior Processes,* **11,** 565–575.

Rachlin, H., Logue, A. W., Gibbon, J., & Frankel, M. (1986). Cognition and behavior in studies of choice. *Psychological Review,* **93,** 33–45.

Rider, D. P. (1983). Preference for mixed versus constant delays of reinforcement: Effect of probability of the short, mixed delay. *Journal of the Experimental Analysis of Behavior,* **39,** 257–266.

Young, J. S. (1981). Discrete-trial choice in pigeons: Effects of reinforcer magnitude. *Journal of the Experimental Analysis of Behavior,* **35,** 23–29.

Received January 25, 1988
Final acceptance July 17, 1988