Variable DRL¹

FRANK A. LOGAN
UNIVERSITY OF NEW MEXICO

Rats were trained on a schedule in which two DRL values were mixed unpredictably. When the two values were quite different, the rats learned to maximize reward by responding first after the shorter interval and then waiting the longer interval if indicated. When the two values were similar, the rats did not maximize reward by always waiting out the longer interval and instead showed predominant adjustment only to the shorter interval.

In the familiar differential reinforcement of low rate (DRL) schedule (e.g., Ferster & Skinner, 1957), a response is rewarded only if preceded by a specified interresponse time (IRT). In a variable differential reinforcement of low rate (VDRL) schedule, the requisite IRT is not constant but assumes one of several values according to some probability distribution. In the particular instance to be described, two equally likely values (S for shorter and L for longer) were mixed with the further restriction that, once one value was programmed, it had to be satisfied before the next scheduled value came into force.

The optimal strategy for maximizing total number of rewards per unit time on this schedule varies with the value of S relative to L. One strategy is to ignore S and treat the condition as DRL-L; the interreinforcement interval will then be L. The alternative strategy is to respond following a reward after S; this response has a 50% chance of being rewarded and, if not rewarded, it will provide a cue that L is in force. The interreinforcement interval will thus be S half the time and S+L the other half the time or, on the average, $S+L/2$. Accordingly, the latter strategy is superior only when $L > 2S$.

The purpose of the present study was to observe the performance of rats when two DRL values were mixed, especially with reference to the strategies outlined above.

Subjects

Ss were 12 Charles River male albino rats, 90 days of age at the beginning of training. They were housed individually with water freely available, and were fed 12 g Purina laboratory chow daily immediately after being removed from the apparatus.

Apparatus

The apparatus was a Foringer operant conditioning box with associated commercial programming equipment. Responses separated by less than 1/2 sec were automatically eliminated. A variety of clicks from this and other apparatuses in the room were audible to the rats, but none was distinctive when reward was set up and frequent observation failed to indicate that any rat was responding to auditory stimuli.

Procedure

During the first exposure to the apparatus, the bar was set for continuous reinforcement, and a pellet was delivered regularly every 60 sec unless a bar press occurred within that time. The few rats that did not learn to respond during this session were given manual shaping, and then all rats were placed directly on the assigned VDRL. Values of S were 5, 10, 15, or 20 sec for groups of three rats; L was 30 sec for all rats.

Each rat had at least 50 1-1/2 h sessions even though no noticeable changes in the records occurred after about 30 sessions. IRTs obtained on two days near the end of training were combined for presentation.

Results

Asymptotic performance can best be characterized by distributions of IRTs drawn separately for the first, the second, and the third-or-later responses following each reward. These latter two distributions are, of course, following nonreward. The essential features of these distributions were remarkably homogeneous for the rats within each group, and hence they are combined in Fig. 1.

Considering first the performance of the 5-30 VDRL rats, it is clear that they successfully learned the two distinctive IRT values and approximated the optimal strategy for maximizing reward. That is to say, their predominant IRT following reward was slightly longer than S; although there was a strong tendency to produce some very short IRTs if that

Fig. 1. Percent frequency of different IRTs after extensive training on the indicated values of a VDRL schedule. The distributions have been separated depending on whether the preceding response was rewarded or not in order to reveal the rats' approximation of the optimal strategy.

response was nonrewarded, they shifted rapidly to IRTs approximating L. Comparable adjustment was achieved by the 10-30 VDRL rats although there was a somewhat greater incidence of long IRTs immediately following reward. In both cases where $L > 2S$, all rats revealed some learning of both DRL values and some separation of these based upon the preceding cues of reward and nonreward.

This was not the case where $L < 2S$. None of the rats in the 15-30 or 20-30 VDRL conditions learned to ignore the shorter interval so as to maximize the density of reinforcement. On the contrary, their performance following reward was reasonably well adjusted to S, while their performance following nonreward displayed considerable variability including many IRTs below S and between S and L, as well as those longer than L that were reinforced.

Discussion

Clearly, the predominant mode of adjustment to a VDRL schedule is to learn to delay responding for the shortest interval of time for which reward is ever obtainable. When the longer interval is substantially different from the shorter one, i.e., $L > 2S$, then the rats also tend to learn the longer interval and to separate S and L responses in relation to reward appropriately. When, however, the longer interval is relatively close to the shorter one, i.e.,

$L < 2S$, either the rats fail to learn the longer interval or they fail to separate S and L responses in relation to reward/nonreward in such a way that this differentiation can be detected.

One observation is of special interest in this connection. It is well known that rats required to delay responding develop ritualistic behavior (e.g., Wilson & Keller, 1953), and it could be argued that the rat is not timing his response per se but is learning other responses that coincidentally consume the requisite time. Adjustment to two mixed DRL values might therefore be expected to require learning two distinctive rituals. Insofar as could be detected by observation, all rats adopted a single ritual. Those who differentiated S from L simply engaged in this ritual (e.g., biting at the food cup) for different lengths of time. Accordingly, these data indicate that adjustment to a DRL schedule is indeed based upon a timing mechanism and is not a coincidental byproduct of reinforcement for other responses.

References

- FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement*. New York: Appleton-Century-Crofts, 1957.
WILSON, P. M., & KELLER, F. S. On the selective reinforcement of spaced responses. *J. Comp. Physiol. Psychol.* 1953, 46, 190-193.

Note

1. This research was supported in part by a grant from the National Science Foundation.