Noncorrected-nonreward during widely-spaced T-maze probability learning¹

KEITH N. CLAYTON VANDERBILT UNIVERSITY

Abstract

The effectiveness of noncorrected-nonreward in T-maze learning was assessed in four probability learning tasks in which different hypotheses regarding the effect of noncorrected-nonreward yielded different asymptotic predictions. It was found that the effectiveness of the event depends on the particular reward schedule used. Generally, the results support the conclusion that the role of noncorrected-nonreward depends on the S's past training experience even under conditions of widely-spaced training.

Problem

This study sought to investigate, with two separate experiments, the role of noncorrected-nonreward during two-choice learning by rats. Noncorrected-nonreward occurs when the rat enters an unbaited goalbox and is removed from that goal-box rather than being permitted to retrace to the other goal-box. Alternative hypotheses for describing the effect of noncorrected-nonreward are that it (a) decreases, (b) has no effect on, or (c) increases response probability. Empirical support for each hypothesis is available from the experiments of (a) Galanter & Bush (1959) and Greeno (1960), (b) Parducci & Polt (1958), and (c) Frankmann (1959). Each of these studies, however, used massed training or forced trials. The present study avoids the potential hazards arising from these procedures by using widely-spaced training with all trials free.

Experiment I

In Experiment I contingent reinforcement probabilities were selected so that the separate hypotheses made different asymptotic response probability predictions. On each trial any of three reward events could occur with probability π ; E_1 , reward for response A_1 ; E_2 , reward for response A_2 ; or E_0 , noncorrected-nonreward. Since the event probabilities were response-contingent there were six π_{ij} values, where i refers to the response made and j , the event which occurs.

The hypotheses are explicitly stated as follows:

$$\text{Hypothesis (I): } (1) P(A_{i,n+1} | A_{i,n} E_0) = P(A_{i,n})$$

Hypothesis I states that noncorrected-nonreward has no effect on response probability, i.e., is an identity operator. The alternative hypotheses are stated generally in Equation (2),

$$\text{Hypothesis (II-IV): } (2) P(A_{i,n+1} | A_{i,n} E_0) = (1 - J) [P(A_{i,n})] + kJ,$$

which says the probability increases or decreases depending on the value of k . Hypotheses II, III, and IV are that k equals 1, $1/2$, and 0, respectively. With the standard linear model assumptions regarding reward and corrected-nonreward (Estes & Suppes, 1959) Equation (2) can be shown to yield the asymptotic prediction:

$$(3) P(A_{1,\infty}) = \frac{\pi_{21} + \pi_{20} - k\pi_{20}}{1 - \pi_{11} + \pi_{21} + \pi_{20} - k(\pi_{10} + \pi_{20})}$$

which shows how the terminal probability depends on k .

Two separate groups were run in the first experiment. Their event probabilities are given in Table 1. The asymptotic predictions are given in Fig. 1.

Method

Sixty-four male albino rats received one T-maze trial daily while 23-hr. food-deprived. Thirty-one rats were in Group A and 33 in Group B. Approximately half of each group was trained with the left side of the maze designated A_1 . A separate reward schedule was constructed for each rat.

Seven days of prehandling, which included adjustment to the feeding schedule and to the reward pellets, was followed by two days of pretraining, which consisted of two spaced rewarded trials in the maze, exposure to both sides being equated by forcing on the second trial. On each trial of experimental training a wooden door, located just beyond the choice-point, was closed behind the rat after a choice was made. On E_0 events the rat was removed from the empty goal-box 10 sec. later. If the response was to be rewarded, 2-45 mg Noyes food pellets were available and the rat was removed 10 sec. later. On corrected trials the choice-point doors were reopened after 15 sec. confinement in the unbaited goal-box and the rat was left in the maze until 10 sec. after entering the baited goal-box. Training lasted 72 days during which time body weights were monitored and kept constant by minor adjustments in the daily ration.

Results and Discussion

As can be seen from the learning curve presented in Fig. 1 Group A, in general, does not seem to have stabilized by the end of 72 trials although the two predictions that the asymptote would be less than unity have already been surpassed. In fact, it seems clear that the terminal response probability for Group A is very near unity since 22 of the 31 Ss made 12 consecutive responses to the A_1 side during the last 12 trials. Group A performance, therefore, is incompatible with Hypotheses III and IV.

On the other hand, as seen in Fig. 1, for Group B, no learning occurred with respect to response A_1 . Only Hypothesis IV, the hypothesis that probability is decreased by E_0 , predicted this outcome. The performance of the two groups suggests, therefore, that the hypothesis which receives support depends on the particular reward schedule utilized, a result which, strictly speaking, is incompatible with all four hypotheses.

Table 1. Reward Probabilities for Exp. I (π_{ij})

		Group A Event (j)			Group B Event (j)		
		E_1	E_2	E_0	E_1	E_2	E_0
Response (i)	A_1	15/24	0	9/24	6/24	2/24	16/24
	A_2	9/24	6/24	9/24	18/24	6/24	0

Experiment II

While the results of Experiment I question all the assumptions which led to the predicted asymptotes (including, for example, the role of corrected-nonreward) the following discussion will assume the fault lies in Hypotheses I-IV. All the hypotheses assume "independence of path," i. e., that the effect of noncorrect-nonreward is independent of the past experience of the S in the learning situation. However, rejection of this assumption has been suggested by the results of a study on successive discrimination reversal (Clayton, 1962). Analysis of the initial trials of the reversals in that study suggested that while one or two noncorrected-nonrewards had no effect on response probability (Hypothesis I), several consecutive nonrewards produced a probability decrement (Hypothesis IV).

According to this interpretation, then, the critical difference between Groups A and B was the difference between their π_{10} values. For Group A this value was 9/24, for B 16/24. It is clear, therefore, that the average length of the consecutive nonrewards was considerably greater for Group B and was probably sufficient to prevent learning (Hypothesis IV). For Group A, however, the average length of consecutive nonrewards was probably insufficient to reject Hypothesis I.

Experiment II was designed as a direct test of this "lag" hypothesis. Two groups of rats were trained on a non-contingent probability learning task, the two groups differing only in terms of whether E_0 occurred on successive trials. For both groups reward was on the A_1 side on 3/4 of the trials and on the A_2 side on the other 1/4 of the trials, and correction was never permitted. Reward assignments were made in blocks of 20 trials and a different schedule was constructed for each rat. For Group C ($N=26$) reward assignments for each of the 20 trials were independent of assignments on the other trials with the slight restriction that consecutive nonrewards for A_1 could not exceed two. For Group D ($N=26$), however, nonrewards of A_1 responses occurred in blocks of five consecutive trials.

If the "lag" hypothesis were correct, Group C would be expected to maximize, i.e., eventually respond with probability above π and near unity to the A_1 side, since the occasional, separated, nonrewards would not be expected to reduce A_1 response probability. Similar results, with massed practice, have been obtained by Parducci & Polt (1958) and North & McDonald (1959). Group D, on the other hand, should not maximize since the consecutive nonrewards would be expected to reduce A_1 probability.

Method

The procedures of Experiment II were essentially the same as Experiment I with the exception of the event probabilities. Another difference was that the pretraining of Experiment II was much more extensive (26 days) in an attempt to insure an initial response probability near 1/2. Experimental training lasted for 60 days and was followed by 10 days of extinction (neither response rewarded).

Results and Discussion

The terminal acquisition A_1 response probability for both groups was near unity—being 0.92 for the last five trials. Group C's performance, of course, is compatible with the Parducci-Polt results. However, it was not expected that Group D would perform at as high

Fig. 1. Learning curves for Groups A and B.

a level as Group C; nevertheless, the two groups do not differ significantly on the last 15 trials of acquisition. A substantial number of both groups was apparently trapped on the A_1 side—during the last 20 trials of acquisition 16/32 animals of Group C and 15/32 Ss of Group D made 20 consecutive A_1 responses.

The acquisition results are rather equivocal with respect to the "lag" hypothesis. The hypothesis could be incorrect; the effectiveness of noncorrected-nonreward may be independent of whether it occurs on consecutive trials. On the other hand, the number of consecutive nonrewards required to reduce response probability may be greater than five.

The extinction data is more informative, since Group C extinguished more rapidly than Group D, as indicated by a significant groups by trials interaction by repeated measures analysis ($F=4.25$, $df=1/48$, $p<0.05$). This continues to support the idea that the effectiveness of noncorrected-nonreward depends on S's past experience even when training is widely-spaced.

References

- Clayton, K. N. The relative effects of forced reward and forced nonreward during widely spaced successive discrimination reversal. *J. comp. physiol. Psychol.*, 1962, 55, 992-997.
- Estes, W. K., & Suppes, P. Foundation of linear models. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford University Press, 1959. Pp. 137-179.
- Frankmann, R. W. Effects of forced rewarded and nonrewarded trials in T-maze learning. Unpublished doctoral dissertation, Indiana University, 1959.
- Galanter, E., & Bush, R. R. Some T-maze experiments. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford University Press, 1959. Pp. 265-289.
- Greeno, J. G. Extinction effects in T-maze choice behavior. Presented at Midwestern Psychological Association, 1960.
- North, A. J., & McDonald, R. D. Discrimination learning as a function of the probability of reinforcement. *J. comp. physiol. Psychol.*, 1959, 52, 342-344.
- Parducci, A., & Polt, J. Correction vs. noncorrection with changing reinforcement schedules. *J. comp. physiol. Psychol.*, 1958, 51, 492-495.

Note

1. This research was supported by grant MH-05576 from the National Institute of Mental Health, United States Public Health Service.