Reaction time as a function of noncontingent reward magnitude[1]

NEIL A. STILLINGS, GORDON A. ALLEN, AND W. K. ESTES
STANFORD UNIVERSITY

College student Ss were given a series of about 250 trials in a simple, nonchoice, reaction time situation, some pre-assigned monetary reward being credited to their accounts following each response. Each trial began, for control Ss, with a warning light, and, for experimental Ss, with a display of the reward value assigned for the trial. The reward received depended in no way upon any property of the S's response. Results revealed a small but reliable tendency for reaction time of experimental Ss to vary inversely with the magnitude of reward anticipated on any given trial.

In one of the few previous studies of response time in relation to reward magnitude in human learning, Keller, Cole, Burke, & Estes (1965) presented data for a modified paired-associate situation. In their study, Ss learned a 25-item paired-associate list, each item having two alternative responses, to each of which was assigned a particular reward magnitude. Reaction time varied systematically with reward magnitude at both intermediate and terminal stages of learning; but the most orderly functions appeared, not for latency vs the difference between the two reward magnitudes assigned for an item, but rather for latency vs the value of the higher of the two rewards. Further, steepness of the function increased with training. These and other findings led to the interpretation that the effects of rewards upon performance involve two phases. Firstly, associations between stimuli and reward values are learned simply by contiguity. Following this learning, anticipation of reward generates facilitative feedback, directly related in intensity to reward magnitude. One function of this feedback is to increase the probability that the given stimulus will be overtly chosen during any interval of time and thus on the average to reduce reaction time (Estes, in press).

In this study we proposed to obtain independent evidence on this interpretation by determining the relationship between reaction time and reward magnitude in a situation which does not involve choices among alternatives. The principal difference between the experimental paradigm of this study and that of the standard runway experiment with varying reward magnitude is that in the present situation the Ss will not need to learn what reward to expect but rather will have this information supplied by the E.

The general plan of the experiment is as follows. Each trial begins with a *go* signal, to which the S simply responds by operating a single response key and following which the S is credited with the reward value, calibrated in points, assigned for the trial. In the experimental condition of principal interest, Ss will be informed before each trial of the assigned reward value. Thus they will be in much the same position as a rat which, at the start of a runway trial, has learned what reward to expect, and will necessarily receive the given reward once it has responded. Only the time between trial onset and the receipt of reward will be under the S's control. By comparing data for this experimental group with appropriate controls we propose to determine whether response time varies simply as a function of amount of anticipated reward even when the amount of reward is in no way contingent upon any property of the S's behavior.

Subjects

The Ss were 40 Stanford undergraduates about equally divided between freshmen and upper-classmen. Most had participated in experiments before but only one had had any experience with a differential reward task.

Apparatus

Stimulus displays were presented by means of a tachistoscope with the input of the generating program adapted for a reaction time experiment. The response button, which activated a microswitch, was located in the center of a panel in front of the S. A patch of light in the center of the tachistoscope screen served as the warning light and four colored cue lights above the screen served as the go signal. Under conditions calling for information to be given about rewards, point values appeared at the appropriate time on the tachistoscope screen.

Procedure

Under each condition Ss were told that we wished to see how fast they could operate the response button following appearance of the go signal on each trial. Between trials the S rested a forefinger lightly on the response button and was to press the button down as quickly as possible upon appearance of the signal. Under the experimental conditions, the Ss were shown prior to the go signal on each trial the number of points that would be given as reward, whereas under the control condition this information was not given. The possible point values were 0, 20, 40, 60, and 80. The program delivered these on a random schedule subject to the constraint that each value occur five times in every 25 trials. Each se-

Table 1.
**Mean Reaction Time in Milliseconds as a Function of Payoff
at Two Stages of Training.**

	Experimental		Control	
Payoff	Trials 1-27	Trials 28-54	Trials 1-27	Trials 28-54
0	282	260	247	234
20	280	253	251	232
40	278	256	247	231
60	274	249	246	232
80	275	250	249	232

quence generated was run twice, once on an experimental S and once on a control S, even though the control S never saw the point values.

For an experimental S the trial began with the appearance of the point value for the trial on the tachistoscope screen, followed by the onset of the go signal after an interval which was drawn from a uniform distribution running from 1.0 to 2.5 sec. Both the payoff display and the go signal remained on until the S responded, following which both went off. Then the next trial began after a rest interval of 5 sec. For the control S the procedure on a trial was exactly the same except that the numerical display of a reward value was replaced by a patch of light on the tachistoscope screen. Under both conditions, Ss were instructed that the rewards did not depend in any way upon their response on any particular trial, that the amount they would be paid depended upon the number of points they scored during the 40 min experimental session, and thus the faster they reacted on the average the more they could earn. After each 10 min of experimental trials, the S was given a 2 min rest interval during which he was informed as to how many points he had obtained up to that time.

The tape output from the apparatus was run into a computer program which stored latencies together with point values for the 20 Ss of each condition. Then the latencies were averaged for each reward condition (even though for the control Ss the distribution would necessarily be random since the Ss could not distinguish among the different types of trials). Payments of Ss varied between $1.50 and $2.50 each for the session.

Results

The number of trials completed within the 40 min experimental session varied from S to S, with most completing from 55 to 58 trials per payoff con-

dition. Virtually complete data are available for 54 trials on each payoff, so mean values for the first and second 27 trials by each group under each payoff condition have been computed and are presented in Table 1. Perhaps the most conspicuous feature of these data is the large and consistent difference between experimental and control conditions, which remains virtually unchanged as both groups reduce their average reaction times with practice. At the least, this result makes it clear that the experimental Ss attended to the preview displays of the reward magnitudes. Presumably the observed differential reflects some interference between the activity involved in processing information from the preview displays and that required for optimal preparatory adjustments for the key pressing response. The improvement in response speed over the series for both groups is clearly independent of reward magnitude.

The finding of primary theoretical interest is the small but systematic effect of reward magnitude upon response time for the experimental group, increasing slightly in the later trial block, to be contrasted with the absence of any trend for the control Ss. To evaluate reliability of the trend under the experimental condition, response times for the lowest two and for the highest two reward magnitudes were pooled for each S, and a t test for paired measures was computed. The obtained t of 2.48 with 19 df is significant well beyond the .05 level. Tukey's test for a monotonic trend, applied to the same data, yields an F value of 5.78, which for 1/19 df is significant beyond the .05 level. Evidently it may be concluded that a significant trend exists for the experimental group. This finding provides another fragment of support for the proposition that, under suitably simplified conditions, probability of response by a human S is related in a simple way to his current state with respect to anticipation of reward.

References

ESTES, W. K. Reinforcement in human learning. Technical Report No. 125, Nonr. 225 (73), Stanford University, Stanford, California, 1967, 50 pp. (To be published in J. Tapp (Ed.), *Reinforcement*, New York: Academic Press, in press.)

KELLER, L., COLE, M., BURKE, C. J., & ESTES, W. K. Reward and information values of trial outcomes in paired-associate learning. *Psychol. Monogr.*, 1965, 79 (Whole No. 605).