Contiguity and reinforcement factors in multiple-choice verbal learning: Parametric influences

BENJAMIN B. LAHEY and WILLIAM S. VERPLANCK, *University of Tennessee, Knoxville, Tenn. 37919*

A multiple-choice learning task was used in four combinations of long-short ITIs and large-small number of items. In each group, half of the Ss were told right or wrong after each choice and half were given no feedback. All four groups showed similar results: Choice without feedback was sufficient to produce an increase in repetition of choices. Repetition after both right and wrong was lower than in the no feedback condition. Repetition after right was higher than after wrong, which was at the chance level, confirming Thorndike's finding. A significant effect of ITI (.001) and number of items (.05) was found in the no feedback groups, but no significant effects were found on repetition after right or wrong.

In Thorndike's most frequently cited study of "multiple-choice learning" (1934, pp. 278ff), Ss were presented with a series of multiple-choice items each consisting of one Spanish and five English words. They were instructed to choose the correct translation of the Spanish word from the English choices and were told right or wrong after each choice. Thorndike evaluated the effects of right and wrong in this experiment against the probability of choosing a given alternative by chance ($1/\text{number of alternatives}$). This value ($p = 0.20$) was taken as a theoretical baseline: He assumed that the frequency of repetition of a choice on the next trial in the absence of right or wrong would be equal to chance. A measure of repetition was calculated by dividing the number of alternatives that were chosen for the first time on Trials 2, 3, or 4 by the number of those alternatives that were chosen again on at least the succeeding trial.

Finding that the frequency of repetition of choices that were followed by wrong was approximately equal to chance, Thorndike concluded that punishment had no effect on the strength of the choice responses.

Tilton questioned Thorndike's assumed baseline. To establish an empirical baseline, he ran a group of Ss in which E said neither right nor wrong after S's choices. He found that the frequency of repeated choices in the absence of differential feedback was

higher than the chance level assumed by Thorndike, but only slightly so. Using the empirical baseline, he concluded that verbal punishment was only weakly effective; averaging across Ss, one occurrence of wrong reduced the incidence of incorrect responses on the following trial by about 11%.

Thus, Tilton's study provides evidence for the effectiveness of verbal punishment in multiple-choice learning, but his results are still in conflict with most of the recent results on punishment. His results show verbal punishment to be only weakly effective in suppressing responses, whereas strong effects of at least one aversive stimulus (electric shock) have been repeatedly demonstrated with both animals and humans (e.g., Azrin & Holz, 1966).

Thorndike's conclusions were dictated by his theoretical baseline, but Tilton's conclusions did not markedly differ from them. The generality of their conclusions, however, may be questioned because of their use of extreme values for two parameters. Both used a large number of items (Thorndike, 200; Tilton, 50 nonsense-syllable items) and long intertrial intervals (ITI), with 1 day of uncontrolled activity between each presentation of the set of items. The effects of selected values of these parameters are examined in this experiment.

Other studies of multiple-choice learning (Buchwald, 1959, 1962; Buss et al, 1956, 1959; Spence, 1964, 1966) have compared groups told right after correct choices and nothing after incorrect choices with groups that were told only wrong after incorrect choices and with other groups told both right and wrong after correct and incorrect choices, respectively. Buchwald (1969) compared the effects of right, wrong, and nothing in Ss that were told all three. No research has been done to establish the comparability of these various within- and between-S comparisons. The present study follows Thorndike and Tilton in comparing the effects of right and wrong for Ss that are told both with Ss that are given no feedback.

SUBJECTS

Forty-eight undergraduates enrolled in introductory psychology at the University of Tennessee, who reportedly did not "know" German, served as Ss. All Ss fulfilled a course requirement. Ss in groups run on more than 1 day were paid \$1.00

for each daily session as well. Sex was counterbalanced between groups, but Ss were otherwise assigned randomly to groups.

PROCEDURE

The Ss were seated individually in a small sound-attenuated room across a table from E and were presented a booklet of five pages of multiple-choice items (one set of items per page and one page per trial¹). Each item in the set consisted of one German word² paired with five English words, lettered A through E. S was instructed to choose one of the English words for each item (by stating the English word orally and underlining it). All words were selected randomly from dictionaries and were balanced for number of letters and first grammatical usage given in the dictionary. All English words were balanced for first letter, and all German words began with the letter "F." No English word was the translation of any German word. A key by which E determined rights and wrongs was constructed by selecting English alternatives randomly, with an equal number from each position, A through E. Four random orders of the items within each set (i.e., for each trial) were used; these sets were presented in 1-2-3-4-1 order for a total of five trials.

The Ss were equally divided into four groups that differed in number of items (15 or 50) and intertrial interval (massed: five trials run in immediate succession; spaced: five trials run one per day for 5 days). Thus, the four groups were: 15 items-massed (15-M), 15 items-spaced (15-S), 50 items-spaced (50-S), and 50 items-massed (50-M). Spaced trials were run daily, Monday through Friday. As reported by Thorndike, running Ss daily required some minor rescheduling of sessions; that is, a few ITIs were greater or less than 24 h, but the differences did not exceed 2 h. The particular parametric values were chosen to cover a broad range of values less extreme than those used by Thorndike and Tilton.

Each of the four groups was equally subdivided into two subgroups on the basis of feedback. Half of the Ss were told right or wrong according to the key (R-W); the other half (controls) received no feedback (NF). E said both right and wrong in a neutral conversational tone. Since what was said to each S in the R-W groups depended on S's choices, the proportions of rights and wrongs on each trial varied. On Trial 1, the proportion approximated chance, i.e., one-fifth right and four-fifths wrong.

All groups were given the following instructions: "Before you is a set of German and English words. Choose the

Table 1
Relative Frequency of Choices Made for First Time That are Repeated on at Least the Next Trial

Conditions				Choices Made for First Time on Trials 1-4	Choices Made for First Time on Trials 2-4
N Items	tr/ Day	N Days	Group	Per Cent Repetition	Per Cent Repetition
Thorndike: 200	1	5	Exper: R:	—	46
			W:	—	23
			Control NF:	—	—
Tilton: 50	1	5	Exper: R:	—	26
			W:	—	23
			Control NF:	—	34
Lahey: 50 (50-M)	5	1	Exper: R:	37	26
			W:	22	17
			Control NF:	63	59
Lahey: 50 (50-S)	1	5	Exper: R:	34	33
			W:	25	25
			Control NF:	51	38
Lahey: 15 (15-M)	5	1	Exper: R:	44	50
			W:	16	13
			Control NF:	71	51
Lahey: 15 (15-S)	1	5	Exper: R:	53	38
			W:	22	19
			Control NF:	49	43

English word that you think is the correct translation of the German word. Pronounce your choice aloud and underline it each time." Before the presentation of the second set of items (second trial), the Ss were told: "Now we will go through the same words again but in a different order." No instructions were given as to feedback or a desired kind of performance.

RESULTS

Table 1 gives the relative frequency of choices that were repeated on at least the next trial for both the present experiment and for Thorndike and Tilton. Following Thorndike, data are reported only on those choices made for the first time on Trials 2, 3, or 4 (and, therefore, followed by right or wrong for the first time). Thorndike eliminated from his computations all items that were chosen correctly on the first trial because his correct choices were the actual translations of the Spanish words. Because correct choices were selected randomly in the present experiment, data for Trials 1-4 are also presented in Table 1. All statistical analyses were made on this data.

The most striking result seen in Table 1 was the most unexpected. In all four groups, per cent repetition after choice-alone was markedly higher than after right. This result can be found in Tilton's tabled results, but it has not been previously discussed. Not only do Ss repeat unreinforced responses at greater than the chance level assumed by Thorndike, but

they do so more frequently than after right.

Equally striking is the fact that repetition after wrong consistently approximates chance for all four groups. This measure does not differ from 0.20 by more than 5% in any group, confirming Thorndike's conclusion that wrong does not suppress repetition significantly below the chance level.

Following Tilton, a "suppression index" was calculated for each S by subtracting per cent repetition of choices followed by wrong from the mean per cent repetition in the corresponding NF group. The index varied from 26% to 55% for the four groups, indicating large differences between repetition after wrong and choice-alone. The suppression index was significantly influenced by both parameters. The massed-trial condition resulted in a greater suppression index ($F = 47.83$, $p < .001$); the 15-item condition also resulted in a somewhat greater suppression index ($F = 6.36$, $p < .05$), but its influence was restricted to the massed-trials groups. Interaction was not significant at the .05 level ($F = 4.13$).

However, the results of the above analysis bear closer examination. Analysis of the data in Table 1 shows that the parametric influences on the suppression index come primarily from their influence on the "control" (NF) groups. Separate ANOVA for repetition after right, wrong, and choice-alone shows that only the NF groups were affected significantly by the parametric variables. In addition, only the ITI effect was significant at the .05 level, and then only for the NF groups ($F = 7.26$, $p < .05$).

Because Thorndike's method of analysis is insensitive to many properties of the data,³ repetitions were also calculated for each trial. These data are presented in

Fig. 1. Per cent repetition of choices made on previous trial followed by right (R), wrong (W), or nothing (NF) as a function of number of trials.

Fig. 1. All choices (i.e., including those made for more than the first time) were included in the analysis. In spite of the restrictiveness of Thorndike's measures, the results of the two analyses were similar. In all groups, and on all trials, choice-alone resulted in greater (or in one case, equal) repetition than after right. Also, repetition after wrong was considerably lower than after either choice-alone or right. ANOVA shows the curves to be significantly different for all groups (15-S: $F = 11.29$, $p < .01$; 15-M: $F = 48.36$, $p < .001$; 50-S: $F = 43.60$, $p < .001$; 50-M: $F = 71.83$, $p < .001$).

The repetition by trials data in Fig. 1 was analyzed for effects of the parametric variables. Separate ANOVA for right, wrong, and NF showed the NF groups to be strongly affected by ITI ($F = 28.91$, $p < .001$) and unaffected by number of items ($F < 1.00$), but with a significant interaction ($F = 5.74$, $p < .05$). The ITI effect for wrong was significant ($F = 5.36$, $p < .05$), but there was no other significant effects on right or wrong. Like the analysis based on the single-choice data (Table 1), the parametric effects on the suppression index calculated on the data by trials are largely attributable to effects on the NF groups. ANOVA for suppression showed results similar to that for the single-choice data (significant ITI effect, $F = 100.67$, $p < .001$; nonsignificant number of items effect, $F = 2.04$, but a significant interaction, $F = 19.01$, $p < .001$). The small effect of number of items seen in the single-trial data of Table 1 is not significant for the more inclusive measures of Fig. 1, but this measure does show a significant interaction with ITI.

DISCUSSION

There are differences in the results of Thorndike, Tilton, and the present study presented in Table 1, but not all in the expected directions. Comparing the studies on repetition after right, Thorndike's show the greatest repetition, in spite of the much larger number of items he used. Tilton, on

the other hand, found the least repetition. These unexpected differences are not large, however, in view of the lack of reported procedural conditions by Thorndike and Tilton. No information is reported, for example, on the S population, identity of the choice response, identity of E, etc. This experiment cannot be considered to be a replication because of Thorndike's limited report. Repetition after wrong and choice-alone are more comparable for the groups. Moreover, the repetition measures within the present experiment are consistent with respect to one another and to the effects of the parametric variables.

In conclusion, the occurrence of both right and wrong after choices resulted in less repetition of choices than after choice-alone. In all groups, repetition after wrong was at about the chance level, as reported by Thorndike. That is, strong suppression was found in all groups relative to the NF groups, but suppression was only to the chance level. The parametric variables strongly influenced repetition after choice-alone but had weak or no effects on repetition after right or wrong. Most importantly, the present data demonstrate that, given instructions to respond to one of a number of stimuli (to one of the five alternatives), the occurrence of that response (using these procedures and under these conditions) is sufficient to increase the probability that the same response will occur on the next presentation of that item. Moreover, the occurrence of right after the response results in a lesser probability that the response will be repeated. These data suggest that the limited influence of wrong found by Tilton can be attributed to his use of extreme parametric values.

REFERENCES

- AZRIN, N. H., & HOLZ, W. C. Punishment. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application*. New York: Appleton-Century-Crofts, 1966. Pp. 380-447.
- BUCHWALD, A. M. Experimental modifications in the effectiveness of verbal reinforcement combinations. *Journal of Experimental Psychology*, 1959, 57, 351-361.

- BUCHWALD, A. M. Variations in the apparent effects of "right" and "wrong" on subsequent behavior. *Journal of Verbal Learning & Verbal Behavior*, 1962, 1, 71-78.
- BUCHWALD, A. M. The effects of "right" and "wrong" on subsequent behavior. *Psychological Review*, 1969, 76, 132-143.
- BUSS, A. H., BRADEN, W., ORGEL, A., & BUSS, E. H. Acquisition and extinction with different verbal reinforcement combinations. *Journal of Experimental Psychology*, 1956, 52, 288-295.
- FERGUSON, E. L., & BUSS, A. H. Supplementary report: Acquisition, extinction and counterconditioning with verbal reinforcement combinations. *Journal of Experimental Psychology*, 1959, 58, 94-95.
- SPENCE, J. T. Performance on a four-alternative task under different reinforcement combinations. *Psychonomic Science*, 1964, 1, 241-242.
- SPENCE, J. T. The effects of verbal reinforcement combination on the performance of a four-alternative discrimination task. *Journal of Verbal Learning & Verbal Behavior*, 1966, 5, 421-428.
- THORNDIKE, E. L. *Fundamentals of learning*. New York: Bureau of Publications, Teachers College, University of Columbia, 1934.
- TILTON, J. W. The effect of right and wrong upon the learning of nonsense syllables in multiple-choice arrangement. *Journal of Educational Psychology*, 1939, 30, 95-115.

NOTES

1. A trial is defined in this case as one presentation of the full set of items.
2. German was used instead of Spanish because of the large number of Tennessee Ss who had been exposed to Spanish.
3. Thorndike's analysis was based solely on those instances in which one alternative of an item is chosen for the first time and is repeated or at least the next trial. In the hypothetical cases below (two sets of three items are shown with choices for five successive trials), a single repetition is counted on Trial 1 for each of the first three examples, and only on Trial 1.

(1)	A	A	A	A	A
(2)	A	A	B	A	A
(3)	A	A	C	D	E

Similarly, no repetitions are scored on any trial of the following cases:

(4)	A	B	A	A	A
(5)	A	B	A	B	A
(6)	A	B	C	D	E

Much information is lost in scoring these items in this way.