The stubborn-error effect in verbal discrimination learning

YUNG CHE KIM, MELVIN H. MARX, and JAMES W. BROYLES
University of Missouri, Columbia, Missouri 65211

The stubborn-error effect (SEE), a tendency toward increasing probability of repetition of specific errors over successive trials, has been reported in tasks using verbal learning materials. The present experiment was primarily designed to investigate the time of stimulus presentation and feedback as determinants of the SEE. Verbal discrimination learning data collected from 80 introductory psychology students confirmed the SEE in a large number of subjects. Its dependence upon temporal factors, most notably the time of stimulus presentation, was also demonstrated. The present data did not support the earlier finding of a gender difference (males more likely than females to contribute to the effect). Finally, the subjects who did show some SEE were found to be generally less capable, as measured on a variety of independently obtained indicators of academic achievement.

Increments in the probability of repeating a prior error in human multiple-choice verbal learning have been shown to be an increasing function of the number of successive repetitions of that same error (Marx, 1971; Marx & Marx, 1980a; Marx, Witter, & Mueller, 1972).

The present experiment was designed primarily to examine this stubborn-error effect (SEE) as a function of the presentation time of the stimuli (four words) and of the feedback (right or wrong signal). Secondary purposes were to assess the role in the SEE of prior associative strength of the verbal cues and to determine the relationship of the SEE to gender and to some measures of academic achievement.

METHOD

Subjects

The subjects were 22 males and 58 females from the introductory psychology course at the University of Missouri–Columbia. They were given credit for participating in the experiment as partial fulfillment of a course requirement.

Design and Materials

The experiment had a 2 by 2 by 2 by 2 factorial design; the manipulated variables were time of word-set presentation (1 or 2 sec), time of feedback (1- or 2-sec presentations of the correct answer), preexperimental association strength, and gender. Preexperimental association strength was manipulated by selection of psychologically familiar words (as judged by 19 psychol-

ogy faculty members who answered a questionnaire), presumed to be relatively "strong" in association values for introductory psychology students nearing the end of their one-semester course. One such word was included in each set of four stimulus words; the other three words, bearing no special relationship to psychology, were matched in terms of word length and Thorndike-Lorge frequency scores.

Three of the four-word sets were used as primacy buffers, and three were used as recency buffers; these were not scored. The intermediate 20 sets were analyzed. Selection of the one "correct" word in each set was random, except that 10 of the correct words were from the 20 "psychology" words (e.g., stimulus, emotion, motive) and the other 10 were from the control words (e.g., silver, equipment, miracle). All of the words in any one set began with the same letter.

Procedure

The lists of 26 sets of four words were presented by slide projector (Kodak Ektagraphic, Model AF-1) for six trials. Each word set was shown for 1 or 2 sec. Subjects were instructed to guess one of the four words in each set on the first trial and to try to remember correct answers and repeat them on subsequent trials. They were allowed 3 sec for response. Responses were made by circling the item number of selection in each set on the answer sheet provided. The correct word was then presented for either 1 or 2 sec. Immediately thereafter, the next word set was shown. A 2-min free recall retention test for correct answers was administered at the end of the sixth trial. The number of subjects in each condition is indicated in Table 1.

RESULTS

The major analysis of these data involved strings of successive errors from Trial 2 on; Trial 1 responses were not scored because they were, necessarily, guesses.

With data of all 80 subjects included, the following percentages of successive error repetition over trials were obtained: 34%, 56%, 72%, and 32%, for Trials 3-6, respectively. This ascending curve, through Trials 3-5, essentially replicates the results obtained earlier in similar situations in which data from all of the subjects were analyzed (e.g., Marx, 1971; Marx et al., 1972).

In the last report on the SEE (Marx & Marx, 1980a),

Table 1
Mean Number of Repeated Errors for R and NR Groups Over Trials

| | Group R | | | | | | | | | | | |
| | Stimulus-Presentation Time | | Feedback Time | | Gender | | R Total (N = 42) | | Group NR (N = 38) | | All Subjects (N = 80) | |
Trial	1 Sec (N = 26)	2 Sec (N = 16)	1 Sec (N = 27)	2 Sec (N = 15)	Male (N = 12)	Female (N = 30)	Mean	PRE	Mean	PRE	Mean	PRE
2	7.61	6.03	7.51	6.14	7.25	7.17	7.19	37	3.71	35	5.54	36
3	3.20	2.73	2.99	2.94	2.83	3.10	3.02	42	.64	17	1.89	34
4	1.95	1.90	2.07	1.78	1.75	2.10	2.00	66			1.05	56
5	1.18	1.05	1.34	.89	.92	1.33	1.45	73			.76	72
6	.66	.15	.59	.22	.30	.52	.46	32			.24	32

Note—PRE = percentage of repetition errors from the last trial.

subjects were differentiated on the basis of their own repetitions, and only one-fifth of them were calculated to have contributed to the effect. A somewhat similar separation of subjects into two groups was made in this study. All subjects who repeated the same error at least on three successive trials were categorized as repeaters (R), and all failing to show such an error string on any word set were called nonrepeaters (NR). This distinction divided the 80 subjects almost exactly in half, with 42 R and 38 NR subjects.

Table 1 provides the mean numbers of successive-error repetitions by trials as a function of stimulus-presentation time, feedback time, and gender. Although there is some suggestion in these data that greater numbers of repeated errors are associated with the shorter (1-sec) times, particularly for the stimulus-presentation variable, no reliable effects were found by analysis of variance; the more sensitive repeated-measures design was not applied because of the limited numbers of subjects and unequal cell sizes.

An additional 2 by 2 analysis of variance, performed with the R vs. NR categorization as the criterion, indicated a reliable effect for the stimulus-presentation variable [$F(1,76) = 4.96$, $p < .05$]; the mean of the 1-sec group was .63, and that of the 2-sec group was .39. In this analysis, the value of 1 was assigned to each of the subjects included in the R group and 0 to each of the subjects in the NR group. However, neither feedback time nor its interaction with presentation time proved reliable.

A separate 2 by 2 analysis of variance was performed on the total number of correct words recalled in the retention test. Again, the stimulus-presentation variable proved reliable (means of 6.82 correctly recalled for the 1-sec group and 9.26 for the 2-sec group) [$F(1,76) = 15.84$, $p < .01$]. Furthermore, although the feedback interval itself was not a reliable effect, it did interact reliably with the stimulus-presentation variable [$F(1,76) = 5.53$, $p < .05$]. This interaction is plotted in Figure 1.

Some analyses of the data incidental to the investigation of the SEE also yielded results of interest. For example, when incorrect guesses occurred on Trial 1, reliably more correct responses were made on Trial 2 after 2-sec stimulus presentation than after 1-sec stimu-

lus presentation [$F(1,76) = 15.53$, $p < .01$] and after 2-sec feedback than after 1-sec feedback [$F(1,76) = 4.93$, $p < .05$]. The percentages of correct responses were 58% and 73% for 1- and 2-sec stimulus presentation and 62% and 70% for 1- and 2-sec feedback. The interaction was marginally reliable [$F(1,76) = 3.07$, $p < .09$].

Further analysis of the two 2 by 2 analyses of variance showed that only the stimulus-presentation variable was reliably associated with the SEE, irrespective of the wrong selections or right selections on Trial 1. When wrong selections were made on Trial 1, the percentages of making the SEE were 15% and 9%, respectively, for 1-sec and 2-sec stimulus presentation [$F(1,76) = 7.81$, $p < .01$]. Similarly, the percentages of making SEE, respectively, for 1-sec and 2-sec stimulus presentation were 11% and 5% when right selections were made on Trial 1 [$F(1,76) = 4.46$, $p < .05$]. The feedback variable and the interaction effect of Stimulus Presentation by Feedback Interval were not reliable for either right or wrong selections on Trial 1.

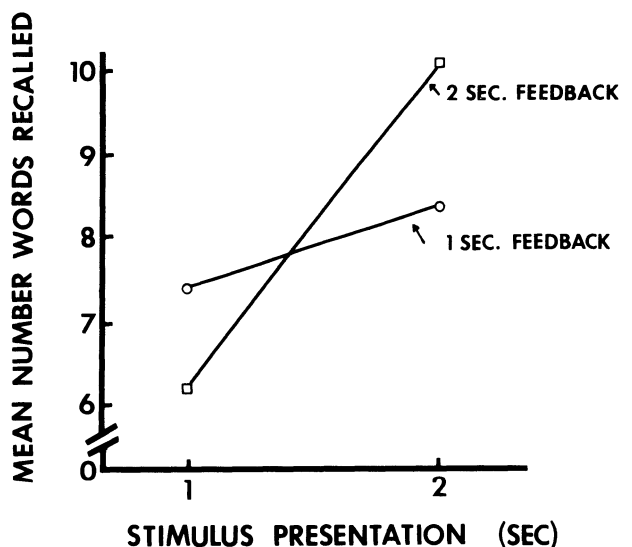Several analyses were performed on the preexperi-

Figure 1. Mean number of correct words recalled for 1- and 2-sec feedback groups as a function of 1- or 2-sec stimulus-presentation time.

mental strength variable ("psychology" words vs. control words). There was no difference in erroneous selection of psychology words vs. control words in guesses on Trial 1, with mean values of 12.52 and 11.73, respectively (t = .95) (with scores adjusted to take account of the 3:1 control-to-psychology word frequency). However, the erroneous selection of psychology words on the second trial was 9.86, compared with a score of 5.84 for control words, and this difference was reliable (t = 2.60, p < .05). The same comparison made on the remaining trials pooled together (Trials 3-6) revealed no reliable difference (7.65 for psychology and 5.58 for control words). But the analysis of the results of the free recall test, given as the final stage of the experiment, showed an opposite, and reliable, difference: More correct control than psychology words (means of 4.78 and 3.24, respectively) were recalled (t = 7.61, p < .01).

A separate one-way analysis of variance of the R data revealed no gender difference. Moreover, 12 of the 22 males (55%) were categorized as Rs, as were 30 of the 58 females (52%). A similarly negligible, and statistically unreliable, difference was found for number of correct words recalled (means of 7.33 for males, 8.29 for females).

Finally, the SEE was related to learning efficiency, as well as to various measures of academic achievement that were already available for introductory psychology students. The index of learning efficiency was taken from the number of correct responses on the second trial, which was, actually, the first trial after feedback had been provided. The subjects whose number of correct responses on the second trial was 13 or less were categorized as low-efficiency learners (L), and the others were categorized as high-efficiency learners (H); the mean number of correct responses on the second trial, for all subjects, was 13.06, and the median was 13. This system of categorization provided an equal number of subjects, 40, in each category.

Contingency analysis of R vs. NR and L vs. H categorization showed that the NR subjects were more efficient learners than were the R subjects [$\chi^2(1) =$ 11.57, p < .01] (the ratio of L vs. H was 27:15 in R group and 11:27 in NR group). It also revealed a relatively high degree of association between the two variables of R vs. NR and L vs. H (C = .36).

The measures of academic achievement that were related to the SEE were high school rank, quantitative, verbal, and total scores on the School and College Aptitude Test (SCAT), and the Missouri Mathematics Placement Test (MMPT). R subjects were compared with NR subjects on these measures, with the values converted to t scores. The NR subjects averaged higher scores on all of these measures, the advantages being 8.18, 5.27, 3.93, 5.42, and 4.46 for high school rank, SCAT verbal, quantitative, and total, and MMPT, respectively. The differences were reliable for high school rank (t = 3.26, p < .01), SCAT verbal (t = 2.10, p < .05), and SCAT

total (t = 2.27, p < .05). These results were confirmed by biserial correlations obtained between categorization into NR or R group and the various achievement scores (rs = .44, .30, .23, .32, and .24 for high school rank, SCAT verbal, SCAT quantitative, SCAT total, and MMPT, respectively; all ps < .01).

DISCUSSION

The present results provide, overall, a confirmation of those reported earlier (Marx, 1971, Marx & Marx, 1980a; Marx et al., 1972). The SEE as a group phenomenon seems to be rather firmly established for verbal learning materials at least. Moreover, the contribution of at least half of the subjects in the present experiment to the effect broadens the base of such contributors, which was estimated at 20% in a previous report (Marx & Marx, 1980a). Of course, the discrepancy in these estimates is obviously due to the different criteria used for categorizing repeaters—only one error string of at least three in the present study, compared with three such strings in the earlier work.

Some determinants of the SEE are suggested by the analysis of its relationships to test scores. Generally, it is clear that subjects with relatively poor academic records are more likely to repeat the same errors more often, as well as to produce a smaller number of correct responses. The extent to which the complexly interacting factors that produce errors in general are related to any special tendency to repeat specific errors (the SEE) will need to be worked out on the basis of research designs tailored to this problem. The statistically reliable relationship of the SEE to the verbal but not the quantitative component of the SCAT in our first analysis was to be expected in light of the verbal character of the materials, but it is interesting that the more comprehensive academic achievement measure of high school rank (as well as the total SCAT score) was also reliably related to the SEE.

The present data offer no support to the earlier finding that males are more likely than females to contribute to the SEE. Further checks on the gender variable will need to control for verbal aptitude and/or achievement, because of the typical female superiority in this respect.

Holding (1970, pp. 727-728) has suggested that stubborn errors of the kind with which we are concerned are either due to the early "learning of mistakes" or to the "formation of incorrect hypotheses or strategies." Although it is very doubtful that the latter process is significantly involved in rote learning of the kind used, it cannot of course be disregarded. Again, more analytic research is needed to decide between these two interpretations, and any others that may be advanced. It should be added that, however such stubborn errors are generated, Kay's (1951) comment that they must be unlearned for mastery is quite obviously applicable; in all of our error-repetition curves, there is some indication of a terminal decline (cf. Table 1 for the present data).

A novel element of the present results is their demonstration of the dependence of the SEE upon temporal factors, most notably the time of stimulus presentation. Forcing subjects to respond quickly clearly encouraged error repetition. More parametric analysis of this important variable is indicated.

With respect to the retention scores, the interesting interaction plotted in Figure 1 suggests the dependence of feedback time on the prior stimulus presentation time. Allowing subjects more time to register correct answers improved their recall of such correct answers only when subjects also had more time to study alternatives.

The results involving the presumed manipulation of associative strength by including both psychology and nonpsychology words are also interesting, but they are more difficult to interpret. Apparently, the manipulation was at least partially effec-

tive, because subjects chose reliably more incorrect psychology than nonpsychology words on the second trial. However, the subsequent reversal of this trend, in the recall data, raises interesting questions. From a cognitive-strategy point of view, it may be hypothesized that some kind of inhibitory process developed over trials, with respect to the selection of psychology words, and was also expressed in the final retention test. Testing of this and the numerous alternative hypotheses that can be generated to explain this result, assuming it is replicable, will of course require carefully designed research. The importance of the associative-strength variable, even within a cognitive-theoretical framework, has been recently suggested (Marx & Marx, 1980b).

REFERENCES

HOLDING, D. H. Repeated errors in motor learning. *Ergonomics,* 1970, 13, 727-734.

KAY, H. Learning of a serial task by different age groups. *Quarterly Journal of Experimental Psychology,* 1951, 3, 166-183.

MARX, M. H. Increased probability of error repetition as a function of number of successive prior repetitions. *Perceptual and Motor Skills,* 1971, 32, 544-546.

MARX, M. H., & MARX, K. Confirmation of the stubborn-error effect in human multiple-choice verbal learning. *Bulletin of the Psychonomic Society,* 1980, 16, 477-479. (a)

MARX, M. H., & MARX, K. Repetition of errors in learning and memory as a function of their prior associative strength. *Bulletin of the Psychonomic Society,* 1980, 16, 435-438. (b)

MARX, M. H., WITTER, D. W., & MUELLER, J. H. Additional data on probability of error repetition following varying numbers of successive prior repetitions. *Perceptual and Motor Skills,* 1972, 34, 525-526.