

## A STUDY OF INCIDENTAL LEARNING\*

*New York University and Barnard College of Columbia University*

M. E. TRESSELT AND M. S. MAYZNER<sup>1</sup>

### A. PROBLEM

Incidental learning has been defined as "learning which apparently takes place without a specific motive or a specific formal instruction and set to learn the activity or material in question" (1, p. 210). Thus, in the incidental learning of verbal material, typically *Ss* are instructed to react to or operate on certain dimensions of the stimulus material and are then tested for retention on other dimensions. For example, Myers (2) asked *Ss* to count all letter *O*'s distributed among a group of other letters and then measured the recall of all other letters present. Most recently the problem of incidental learning has been subjected to an intensive analysis by Postman and his associates (3, 4, 5), employing a variety of techniques for the presentation of the stimulus material to the intentional and incidental learning groups.

The present study takes as its point of departure certain of the theoretical notions advanced by Postman and his associates to account for some of the phenomena of incidental learning. Thus, Postman has employed as an orienting hypothesis: that incidental learning is a positive function of the number of differential responses evoked by the stimulus materials.

This study attempts to manipulate the number of differential responses evoked by the stimulus by varying the operations to be performed on the stimulus by *S*. Specifically, *Ss* are presented with a list of words and instructed to perform certain operations on the words, i. e., one group is instructed to cross-out all vowels, another group to copy the list, and a third group is instructed to judge the words for inclusion in the concept "economic." Following completion of these tasks, *Ss* are tested for their retention of the word-list. Interest is focused primarily on the effects on retention of the different operations performed on the words. In particular, the different operations may be ordered in terms of the degree to which *S* engages the stimulus word qua word. That is, at one extreme where *S* is instructed to cross-out vowels, the stimulus word as such, may be treated merely as a

\* Received in the Editorial Office on August 27, 1960, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This study was supported in part by a grant made to M. E. Tresselt by the New York University Arts and Science Research Fund.

series of letters which *S* searches for vowels, while with the judgment operation, an explicit differential response to the word itself and its associated meanings is required. Perhaps somewhere between these two extremes lies the operation of copying the word. It seems plausible to assume then that as the degree of *S*'s engagement with the stimulus words increases, there will be a corresponding increase in the number of differential responses evoked by these words, and therefore a greater degree of incidental learning should occur.

## B. METHOD

### 1. *Subjects*

Two hundred *Ss* obtained from undergraduate psychology courses at New York University were employed in the study.

### 2. *Materials*

The stimulus material presented to *Ss* consisted of a list of 100 words (shown in Table 1). These 100 words were selected from a larger sample of 600 words employed in an earlier scaling study (6). In this earlier scaling study, 100 *Ss* were asked to judge or check each of 600 words as belonging to none, 1, 2, 3, 4, 5, or 6 concepts, the concepts being theoretical, economic, aesthetic, social, political, and religious. By summing the number of checks for each word for the 100 *Ss*, 600 word-profiles were obtained, which showed the frequency with which each word was judged as belonging to each concept. Using only the scale values (i. e., the number of checks for a given concept) for the economic concept, 100 words were selected for this study, so that the words appeared as nearly as possible at equal intervals along the scale of checkability. That is, the words in Table 1 are arranged in order of increasing checkability (as obtained in the earlier scaling study) and the associated scale values (shown under Column I) range from 1 to 100, in equal steps of one unit, except in a few cases, where there is a two- or three-unit jump.

### 3. *Procedure*

The *Ss* were presented with the words shown in Table 1, completely randomized as to initial scale value, on mimeographed sheets. Of the 200 *Ss* tested, 100 *Ss* were instructed to judge the degree to which each word belonged or could be included in the concept "economic," using a seven-point scale. The number "one" was to be used if they judged the word as not belonging or belonging to a very small degree to the concept; the number "seven" was to be used if they judged the word as belonging completely or to a

TABLE 1  
WORDS AND ASSOCIATED SCALE VALUES USED IN THE STUDY

Words	Scale values		Words	Scale values	
	I*	II**M		I*	II**M
1. Poem	1	1.17	51. Liquidate	51	4.57
2. Horizon	2	1.86	52. Partner	52	4.84
3. Priest	3	1.43	53. Volume	53	4.61
4. Kick	4	1.21	54. Union	54	5.70
5. Blend	5	2.16	55. Power	55	5.17
6. Skeptic	6	2.08	56. Report	56	3.67
7. Classical	7	1.57	57. Policy	57	5.14
8. Loyalty	8	2.34	58. Quality	58	4.61
9. Counterpoint	9	2.09	59. Personnel	59	5.20
10. King	10	2.68	60. Save	59	5.06
11. Banquet	11	2.30	61. Career	61	4.65
12. Violent	12	1.73	62. Ransom	62	2.91
13. Illumination	13	2.29	63. System	63	5.43
14. Investigate	14	3.48	64. Utilize	64	5.02
15. Civil	15	3.33	65. Reward	65	3.43
16. Appeal	16	2.80	66. Boss	65	4.99
17. Execute	17	2.85	67. Statistics	67	5.41
18. Evolution	18	3.01	68. Work	68	6.03
19. Judgment	19	4.05	69. Value	69	5.59
20. Tea	20	1.97	70. Office	71	4.37
21. Trunk	21	1.28	71. Bonds	71	6.42
22. Comparison	22	3.02	72. Trust	72	5.66
23. Chair	23	1.53	73. Department	73	3.45
24. Form	24	2.53	74. Increase	74	4.20
25. Travel	25	3.27	75. Client	75	4.54
26. Method	26	4.16	76. Demand	76	5.40
27. Selfish	27	2.94	77. Appropriation	77	5.66
28. Entry	28	3.28	78. Competition	78	6.12
29. Win	29	3.05	79. Margin	79	5.22
30. Blanket	30	1.64	80. Cartel	80	5.54
31. College	31	3.98	81. Consumption	82	5.87
32. Object	32	2.34	82. Machinery	83	4.95
33. Drive	33	3.89	83. Petroleum	83	4.66
34. Experience	34	3.62	84. Coal	84	4.07
35. Free	35	3.46	85. Rich	85	5.10
36. Society	36	5.13	86. Imports	86	6.30
37. Restriction	37	3.97	87. Management	87	5.92
38. Concrete	38	3.14	88. Rates	89	5.78
39. Socialism	39	4.56	89. Receipt	89	5.10
40. Standards	40	5.33	90. Budget	92	6.44
41. Charity	41	3.26	91. Expenditure	92	5.85
42. Bridge	42	2.90	92. Sale	92	5.80
43. Area	43	2.95	93. Monopoly	93	6.40
44. Company	44	5.37	94. Depreciation	94	5.58
45. Balance	45	5.17	95. Invest	95	6.54
46. Flood	47	3.41	96. Trade	96	6.36
47. Organize	47	5.46	97. Assets	96	6.20
48. Wear	48	2.23	98. Finance	98	6.67
49. Deal	49	4.62	99. Wages	99	6.35
50. News	50	3.60	100. Industry	100	6.32

\* Column I gives the scale values (frequency of checkability) for the words from the earlier scaling study.

\*\* Column II gives the scale values for the words obtained in the present study.

very strong degree to the concept. The intermediate numbers "two" through "six" were to be used to express intervening degrees of belonging or inclusion in the concept. The results of these judgments are given under Column II in Table 1. The scale values shown were calculated by averaging the ratings (1 through 7) given by the 100 Ss for each word. Immediately after the judging task, 50 of the 100 Ss were asked to recall as many of the 100 words as possible, while the remaining 50 Ss were asked for their recall performance approximately 48 hours after the initial judging task. All Ss were given 10 minutes for their recall task, and had no prior knowledge that they would be required to recall the words they had judged. The remaining 100 Ss were divided into two groups of 50 Ss each. One group was instructed to cross-out all the vowels in each word—Ss were told *Es* were interested in developing a perceptual test. The other group was instructed to copy each word—Ss were told *Es* were interested in obtaining samples of handwriting. Both were then immediately tested for their recall of the words. Ten minutes was again allowed for the recall test, of which Ss had no prior knowledge. All testing was carried out during regular class periods, which allowed ample time for the initial and recall tasks. None of the Ss when questioned expressed any suspicion that a subsequent recall test was planned.

### C. RESULTS AND DISCUSSION

The principal results of the study are presented in Table 2, which shows the means and standard deviations for the number of words recalled cor-

TABLE 2  
MEANS AND STANDARD DEVIATIONS FOR THE NUMBER OF WORDS RECALLED CORRECTLY AND INCORRECTLY FOR THE FOUR CONDITIONS—VOWELS (V), HANDWRITING (H), IMMEDIATE (I), AND 48 HOUR DELAY (48)—OF THE STUDY.

		V	H	I	48
Correct	Mean	4.30	8.38	16.86	6.60
	SD	3.00	3.92	7.20	3.20
Incorrect	Mean	14.98	6.34	5.68	12.56
	SD	12.16	5.00	4.22	6.80

rectly and incorrectly for the four conditions of the study, i. e., a crossing-out vowels task (V), a handwriting task (H), and a judgment task with immediate (I) and 48 hour recall (48). The 48 condition, while not directly relevant to the major hypothesis of the study concerning the effect of differential responses, was included to provide some supplementary data on the effect of a moderate time interval between exposure to the stimulus material and its subsequent recall.

It was predicted that as the stimulus material evoked increasing numbers of differential responses, from the V to the H to the I tasks, there should occur a concomitant increase in the recall scores, and inspection of Table 2 indicates such an increase, from a mean of 4.30 words recalled correctly in the V task to a mean of 8.38 words recalled correctly in the H task, to a mean of 16.86 words recalled correctly in the I task. A simple analysis of variance for the V, H, and I tasks yields an  $F = 78$ , which is highly significant ( $P < .001$ ) and comparisons of V with H, H with I, and V with I tasks are all significant ( $P < .01$ ). The 48 condition shows a mean of 6.60 words recalled correctly, which is significantly different ( $P < .01$ ) from the mean of 16.86 words recalled correctly in the I condition, indicating as might be expected that a delay of 48 hours, from exposure of the stimulus material to its subsequent recall, can produce a marked and significant decrement in recall performance.

With respect to the mean number of words recalled incorrectly, for the V, H, and I tasks, rather than an increase in scores from the V to the I tasks, there is a highly significant decrease (simple analysis of variance yields an  $F = 20$ ,  $P < .001$ ), from a mean of 14.98 words recalled incorrectly in the V task, to a mean of 6.34 words recalled incorrectly in the H task, to a mean of 5.68 words recalled incorrectly in the I task. A comparison of the I and 48 tasks, reveals a significant increase ( $P < .01$ ) from a mean of 5.68 words recalled incorrectly in the I task to a mean of 12.56 words recalled incorrectly in the 48 task. Thus, it appears that: (a) as the assumed number of differential responses evoked by the stimulus material increases—from the V to the I tasks—there is a significant increase in the number of words recalled correctly, but a significant decrease in the number of words recalled incorrectly, and (b) a delay of 48 hours from exposure to recall, produces a significant decrease in the number of words recalled correctly, in contrast to immediate recall, but a significant increase in the number of words recalled incorrectly.

The results in Table 2 while clearly supporting the main hypothesis of the study, also raise additional issues that require discussion. It would seem that in the type of incidental recall task that is employed in this study, *S* not only produces responses that appeared on the stimulus word-list (i. e., correct recall), but also produces a varying and relatively large number of responses that did not appear on the stimulus word-list (i. e., incorrect recall). If we examine these "incorrect" responses, we find that their number varies systematically as a function of the conditions under which the stimulus word-list was presented (i. e., the V, H, I, and 48 tasks). In general, this variation

is what might be expected, in that the number of correct responses is inversely related to the number of incorrect responses. That is, when *S* is unable to recall many correct responses, he apparently elaborates with a large number of incorrect responses, and conversely when *S* is able to recall many correct responses, he greatly reduces the number of elaborated incorrect responses. It almost appears from these results that in the 10 minutes allowed for recall, *S* is set to produce a relatively constant number of total responses, that varies from 14.72 in the H task to 22.54 in the I task, but varying systematically in the proportion of correct and incorrect responses.

To understand further the processes behind the production of incorrect responses, an analysis was made of the content of these responses. In the V condition, which produced a mean of 14.98 incorrect responses, or a total sample for the 50 *Ss* of 749 words, there were 601 different words. That is, the production of words was highly unique for each *S*, and in fact only three words (namely, agree, appear, and believe) out of the 601 words produced, were shared as common responses by five *Ss*. All other words showing commonality were for fewer than five *Ss*. In the H condition, which produced a total sample of 317 words, there were 258 different words, and only one word revealed a commonality as high as five *Ss*. For the I and 48 conditions, the results are somewhat different. In the I condition, which produced a total sample of 284 words, there were 170 different words, nine of which showed commonalities of five *Ss* or higher. These words, with the commonalities given in parenthesis, were as follows: investment (5), inflation (5), price (6), interest (6), export (8), stock (9), market (9), business (9), and money (11). In the 48 condition, which produced a total sample of 628 words, there were 320 different words, 28 of which showed commonalities of five *Ss* or higher, with the highest word "money" showing a commonality of 20 *Ss*.

These results suggest that the process governing the production of incorrect responses is not identical in the four conditions of the study. The V and H conditions give rise to large samples of unrelated, unique responses, whose content such as "agree," "appear," etc., bear little or no relation to the concept "economic." The I and 48 conditions, however, are characterized by a higher degree of commonality among the responses and bear a far more direct relationship to the concept "economic," as evidenced by their content such as "investment," "stock," "money," etc. Since the I and 48 conditions involve explicit judging the words for inclusion in the concept "economic," while the V and H conditions do not involve any such judgments, it is not surprising to find more commonality of responses and responses related to the

economic concept, in the I and 48 conditions, than in the V and H conditions. It might be assumed then that in the I and 48 conditions, the "economic" judgment task sensitizes S's response hierarchy for "economic" content responses, thus affecting the production of incorrect responses in the manner shown.

If S's response hierarchy for "economic" content-type responses is sensitized by the "economic" judgment task, we should find additional evidence of this in the production of correct responses. Specifically, we might expect that if we examine the content of the correct responses most frequently recalled versus least frequently recalled, there should be systematic differences in the content as a function of the four conditions of the study. Since the I and 48 groups judged the words for inclusion in the economic concepts on a seven-point scale, scale values exist which provide a quantitative estimate of the "economic" content of each word. For example, as shown in Table 1, word number 1, "poem" with a scale value of 1.17 may be said to have low economic content, while word number 100, "industry" with a scale value of 6.32 may be said to have high economic content.

In making the analysis, the mean scale value of the 10 most frequently correctly recalled words and the mean scale value of the 10 most infrequently correctly recalled words were compared. For example, in the I condition the 10 most frequently correctly recalled words (frequency or commonality scores are in parenthesis) were: power (32), boss (29), college (28), imports (22), priest (22), cartel (21), coal (21), society (21), wages (20), and bonds (19). The 10 most infrequently correctly recalled words in the I condition were: comparison (1), drive (1), investigate (1), illumination (1), judgment (1), quality (1), restriction (1), volume (1), violent (1), and civil (1). The mean scale value for the 10 most frequent words given above is 4.94, while the mean scale value for the 10 most infrequent words is 3.50, a difference which is significant at less than the .01 level. For the 48 condition, the corresponding mean scale values are 6.03 and 3.15, a difference which is significant at less than the .001 level. In contrast, for the V and H conditions, the corresponding mean scale values are, for V, 3.50 and 3.99, and for H, 4.86 and 4.74, neither difference yielding statistical significance.

The failure of the V and H conditions to yield a significant difference in scale value between the most frequent and most infrequent recalled words, in contrast to the I and 48 conditions, where the difference was highly significant, indicates clearly the sensitizing effect of the "economic" judgment task to raise the response thresholds for words having high economic content.

That is, in the I and 48 conditions, the words most frequently recalled have high economic content, in contrast to the words most infrequently recalled, whereas in the V and H conditions, where no sensitization to "economic" takes place because of the task instructions, there is no difference in economic content between the words most frequently and most infrequently recalled. The sensitizing effect produced by the "economic" judgment task is evidenced further, when we compare the mean scale values for the most frequently recalled words in the I and 48 conditions. As already indicated these values are 4.94 and 6.03, respectively, a difference which is significant at less than the .05 level. Thus, it appears that a delay of 48 hours from exposure to recall, further magnifies the sensitizing effect, as the most frequently recalled words have an even higher economic content (6.03), than when recall is immediate (4.94). This increased potential to produce responses with high economic content, after a 48 hour delay, probably represents a differential weakening of the response strengths for the words in the list, with the response strengths for the words having high economic content being maintained to a greater degree because of the sensitizing effect of the judgment task.

In the results so far discussed, we have noted both a facilitating and sensitizing effect on recall scores for the "economic" judgment task. A serious methodological question must be raised however, as to the role of the "free production" of economic content words, rather than their recall. That is, in the recall test, in both the I and 48 conditions, to what extent is *S* "recalling" words previously exposed to him, and to what extent is *S* "freely producing" words he believes possess economic content? In an attempt to answer this question, an additional group of 25 *Ss* were asked to produce, during a 10-minute period, all the words they felt might be included in the concept "economic." For this sample of 25 *Ss*, the mean number of words produced was 58.20, yielding a total sample of 1455 words, of which 702 were unique responses. A comparison then between the mean number of words produced in 10 minutes (58.20) versus the mean number of words recalled in 10 minutes (16.86), indicates a very significant difference ( $P < .001$ ) between the two conditions. These results imply that if there is any tendency to "free produce" in the recall situation, it represents a drastically reduced type of free production, in contrast to the pure production condition.

Another type of comparison between the free production and recall situations may be made with regard to the commonality of responses in the two situations. If we compare the 20 most frequently produced words in the free



production situation with the 20 most frequently recalled words in the I and 48 conditions, we find for the 48 condition only four of the 20 words identical in both samples, and for the I condition only three of the 20 words identical in both samples. These results, that is, both total number of responses produced and commonality of responses, are suggestive that the processes underlying response production in the free production and recall situations are different, and therefore, in the recall situation, the facilitating and sensitizing effects of the "economic" judgment task are not explainable solely in terms of a tendency of Ss to merely free produce words having economic content, but some active recall process also must be involved.

#### D. SUMMARY

Since Postman has suggested that incidental learning is a positive function of the number of differential responses evoked by the stimulus material, this study was designed to manipulate the number of such differential responses by varying the operations to be performed on the stimulus by S. Specifically, Ss were presented with a list of words and instructed to perform certain operations on the words, i. e., one group was instructed to cross-out all vowels, another group to copy the list, and a third group to judge the words for inclusion in the concept "economic." Following completion of these tasks, which are assumed to evoke an increasing number of differential responses, Ss were tested for their retention of the word-list. The results showed clearly that as the number of differential responses increased, from the vowel to the copying to the judgment tasks, there was a highly significant increase in the recall scores, as expected.

#### REFERENCES

1. MCGEOCH, J. A., & IRION, A. L. *The Psychology of Human Learning*. New York: Longmans, Green, 1952.
2. MYERS, G. C. A. A study in incidental memory. *Arch. of Psychol.*, 1913, 4, No. 26.
3. POSTMAN, L., & PHILLIPS, L. W. Studies in incidental learning: I. The effects of crowding and isolation. *J. Exp. Psychol.*, 1954, 48, 48-56.
4. POSTMAN, L., ADAMS, P. A., & PHILLIPS, L. W. Studies in incidental learning: II. The effects of association value and of the method of testing. *J. Exp. Psychol.*, 1955, 49, 1-10.
5. POSTMAN, L., & ADAMS, P. A. Studies in incidental learning: VIII. The effects of contextual determination. *J. Exp. Psychol.*, 1960, 59, 153-164.
6. PRITCHARD, L., MAYZNER, M. S., & TRESSELT, M. E. The Mayzner-Tresselt word-concept check-list. *Psychol. Newsl.*, 1955, 7, 27-33.

15 Sydney Avenue  
Deal, New Jersey