

Encoding specificity and contextual similarity

MICHAEL L. EPSTEIN, DAVID A. DUPREE, and LEE A. GRONIKOWSKI
Rider College, Lawrenceville, New Jersey 08648

Taxonomic category headings were varied in a 2 by 2 factorial design at input and output in the processing and recall of concrete nouns in two experiments with college undergraduates. Experiment 1 showed that cuing at recall was beneficial regardless of whether cues were present during processing, but was maximally effective if cues were present at both stages. Experiment 2 examined whether the effectiveness of recall cues was due to their reinstating the encoding context at recall. The results supported neither a generation-recognition model of recall nor a strict interpretation of the encoding specificity principle. It is suggested that situationally specific mini-theories be formulated to account for cued recall phenomena.

The encoding specificity principle states that the effectiveness of a retrieval cue is dependent upon its having been encoded in some manner at input with the to-be-remembered (TBR) material. Tulving and his associates (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1973) have shown that subjects asked to study a TBR word and its weak associate recalled the TBR item better when cued with the weak associate than with a strong associate that was not present during input. For example, subjects who studied the word CHAIR in the presence of its weak associate GLUE, were later able to recall CHAIR when cued with GLUE, although they could not recall the TBR word cued with TABLE. Furthermore, if following learning (but prior to recall) subjects were asked to produce associates to the word TABLE and they produced the word CHAIR, they frequently did not recognize it as a word they initially encoded. However, they were still able to recall CHAIR when cued with GLUE.

A recent test of the encoding specificity principle by Epstein and Dupree (1977) did not support the above findings. Epstein and Dupree found that, following an incidental learning task in which subjects sorted some words under category headings while leaving other words uncategorized, recall for the initially uncategorized words was improved by presenting appropriate category heading cues. Recall for these words was equivalent to those cued at both encoding and recall. The authors suggested that some sort of generation-recognition model was more appropriate to account for the results than the encoding specificity principle. According to the former model, subjects in the above study could have generated exemplars of the various categories when given a category cue at recall. Subjects would then recognize from their list of generated items those items they had processed during input.

Tulving (Note 1) questioned the findings of the Epstein-Dupree (1977) study because of the lack of

equal-sized categories in the initially categorized vs. the uncategorized conditions. There was a slight tendency for better recall from smaller categories, and the initially uncategorized conditions in the Epstein-Dupree study consisted of single-item categories.

A number of other researchers favored the generation-recognition model relative to the encoding specificity principle (e.g., Humphreys & Galbraith, 1975; Martin, 1975; Pellegrino & Salzberg, 1975). Santa and Lamwers (1974, 1977) stated that studies showing recognition failure of recallable words are not valid evidence of encoding specificity. Rather, the typical study showing recognition failure of recallable words requires subjects to process an association between a TBR word and its weak associate. Subjects fail to recognize the TBR word when they produce it to a strong associate, not for reasons of encoding specificity, but because they had initially processed the association between the words, not a single item. Thus, following a recognition failure they are still able to produce the TBR word when asked to recall the association between it and its weak associate.

EXPERIMENT 1

The first experiment in the present study essentially replicated the earlier Epstein-Dupree (1977) study, with the exception that category sizes were equated in all the input-output cuing conditions. In addition, the effects of cuing at input and/or output were examined under intentional as well as incidental learning conditions. If the encoding specificity principle is valid, we would expect recall for words given the same cues at input and output to be recalled best and recall for words cued only at output to be no better than for those cued neither at input nor at output.

Method

Subjects. Subjects were 34 (17 male, 17 female) undergraduate students at Rider College. Twenty-two students participated in the incidental learning conditions while 12 learned intentionally. All were fulfilling an introductory psychology course requirement and had no previous experience in psychological experiments.

We thank Richard Deni for his helpful comments on an earlier draft of this paper. David A. Dupree is now at the University of Illinois, Urbana, Illinois 61801.

Design and Procedure. Subjects entered the laboratory in groups of four to eight. They were told that the study involved a card-sorting task in which they were to sort cards under appropriate category headings. Eight 3 x 5 in. index cards were placed on a table in front of each subject; on each card was printed a taxonomic category heading (e.g., TOOLS, AIR-CRAFT, VEGETABLES) in capital letters. A ninth card with the word UNCATEGORIZED on it was also placed in front of the subjects.

Each student was handed a deck of 64 index cards, each with a concrete noun printed on it in lowercase letters. Subjects were instructed to sort the cards as quickly as possible under the appropriate headings. The decks were constructed so that there were two two-item, two three-item, two five-item, and two six-item categories. An additional 32 items were not exemplars of any of the eight categories and were therefore to be placed under the heading UNCATEGORIZED. The 32 uncategorized cards actually consisted of two categories each of 2, 3, 5, and 6 exemplars that were not appropriate to the other eight categories. The card-sorting task was repeated three times, each time with the cards in a different random order. In addition, subjects in the intentional learning condition were told to try to learn the words for a later retention test.

Immediately following the card-sorting task, subjects were given a sheet of paper with 64 spaces on it and asked to recall as many of the words as they could during a 6-min period. When time had elapsed, the free recall sheets were collected and a second sheet of paper was distributed. On the second sheet were 16 category headings, one from each of the different-sized categories that was present during the card-sorting task and appropriate category headings for each of the different-sized categories that had been left uncategorized during processing. An appropriate number of blank spaces was provided under each heading. An additional 32 blank spaces were provided at the bottom of the page for the 32 remaining words left uncued at recall.

It is important to note that 16 of the remaining 32 words had originally been categorized under appropriate headings at input. To summarize, at recall half the items that had been categorized at input were cued with category headings and half were not. In addition, half of the originally UNCATEGORIZED items were cued at recall with appropriate category headings, while the remaining 16 items were cued neither at input nor at recall. Six minutes were allotted for "cued" recall.

The presence of category cues during processing and/or recall resulted in a 2 by 2 within-subjects factorial design. That is, there were four experimental conditions: (1) category headings present during both processing and recall (C-C), (2) category headings during processing but not at recall (C-U), (3) category headings present during neither processing nor recall (U-U). The material was counterbalanced for subjects so that each word occurred in each condition equally often. In all conditions the TBR words were obvious instances of the appropriate taxonomic categories.

Results and Discussion

Table 1 summarizes free recall scores for intentional and incidental learning conditions in Experiment 1. Scores were converted into percentage of items recalled from each category size to allow comparisons between different-sized categories. It is apparent that intentionality of learning had little effect on the level of free recall. Except for better recall of two-item categories in the incidental learning group and six-item categories in the intentional learning group, there was almost no difference due to intentionality of learning. The mean percentage of items recalled that had been categorized during processing in the incidental and

Table 1
Mean Percentage of Free Recall as a Function of Orienting Condition, Cue Presence, and Category Size in Experiment 1

Category Size	Cuing Condition			
	Cued		Uncued	
	Intentional	Incidental	Intentional	Incidental
2	56.25	77.27	27.08	30.68
3	72.22	65.83	30.56	28.79
5	58.33	55.00	36.67	35.91
6	61.11	62.92	29.17	40.91
Overall Mean	61.72	62.78	31.51	35.80

intentional conditions was 62.78% and 61.72%, respectively. For uncategorized items mean recall was 35.80% and 31.51%, respectively. Because of minimal differences due to learning intentionality, the remainder of this section is limited to a discussion of the performance of the incidental learning group to facilitate comparisons to earlier research.

The presence or absence of category cues at input and the four levels of category size produced a 2 by 4 within-subjects factorial design. As in the Epstein-Dupree (1977) study, free recall following input categorization was far superior to those conditions in which items were not categorized [$\bar{X}_{cat} = 62.78\%$, $\bar{X}_{uncat} = 35.80\%$, $F(1,21) = 103.14$, $p < .001$]. The card-sorting task apparently resulted in differential storage for the categorized relative to the uncategorized items; during free recall the former items were more readily available and/or accessible. There was no significant effect of category size upon the percentage of items recalled.

The data from the cued recall conditions were converted into percentage of items recalled per category and analyzed in a 2 by 2 by 4 within-subjects analysis of variance; the factors represent input categorization, output cuing, and category size, respectively. The data are summarized in Table 2. As with the free recall scores, categorization during processing led to better recall [$F(1,21) = 48.66$, $p < .001$]. Prompting subjects with the category headings at recall produced an increase in the percentage of items recalled relative to when prompts were not used [$F(1,21) = 81.06$, $p < .001$]. There was no significant effect of category size on recall. However, category size interacted significantly with categorization during processing and with output cuing [$F(1,21) = 5.92$, $p < .001$; $F(1,21) = 3.94$, $p < .01$, respectively]. That is, the effect of categorization during processing resulted in better recall of items from the smaller categories. The lack of category cues at recall was more detrimental to items from the six-item category than the others. The theoretical significance of these interactions is unclear.

The critical comparison of conditions in testing the encoding specificity principle involves recall in the U-C vs. the U-U conditions. According to the principle, category cues at recall would not be expected to improve

recall unless those cues were also present during processing. Obviously, recall in the former condition was substantially better than in the latter ($\bar{X}_{U-C} = 57.10\%$, $\bar{X}_{U-U} = 26.14\%$). A post hoc test showed the differences were significant [$t(21) = 8.23$, $p < .001$].

While the data do not support the encoding specificity principle, neither can they be construed as support for the generation-recognition model. If subjects used the taxonomic category headings at recall in order to generate exemplars of the category, some of which were then recognized as having been processed during the sorting task, recall in the U-C and C-C conditions should not differ significantly. Instead, recall in the latter condition was substantially better for each sized category [$\bar{X}_{C-C} = 73.58\%$, $\bar{X}_{U-C} = 57.10\%$; $t(21) = 5.70$, $p < .001$]. Even though cuing at recall improved performance, whether or not the category heading was present during processing, cue effectiveness was reduced if the headings were absent during processing.

Tulving (Note 1) conceptually modified the encoding specificity principle in a manner that accommodates the above results. Rather than maintaining the position that a cue is effective at recall only if it was present at input, he holds that cuing at recall is *most* effective if the cue is also present during initial encoding. According to this position, recall in the U-C condition should be poorer than in the C-C condition, but still better than in the C-U or U-U condition. These predictions are supported by the present findings. Of course, a modified generation-recognition model could as easily account for the results by including the assumption that one would be more likely to generate an exemplar when cued with a category heading if that item had also been overtly organized under that heading during processing.

A third possibility may account for the results of Experiment 1: Presenting subjects with cues at recall which were also present during processing may lead to better recall due to the similar context in the processing and recall conditions. The beneficial effects of cuing at recall attributed to encoding specificity would then be due to the more general effect of reestablishing contextual cues at recall (Bilodeau & Schlosberg, 1951). A second experiment was conducted to test this possibility.

EXPERIMENT 2

Wallace (1978) has shown that recognition failure of

recallable words, a major phenomenon used in support of the encoding specificity principle, results from the removal of the original study cues in the recognition testing context and their restoration during cued recall. According to this position, recall should improve when original contextual cues are restored at recall relative to occasions when they are not. We would expect, therefore, that the percentage of items recalled in the various conditions of Experiment 1 would increase as contextual similarity increased by presenting subjects with exemplars of each taxonomic category as additional cues at recall.

Method

Subjects. Subjects were 16 (8 male, 8 female) undergraduate students from the same population as that in Experiment 1.

Design and Procedure. The same basic procedure as that in Experiment 1 was used. The critical difference between the experiments occurred during cued recall in Experiment 2. An exemplar from each category was provided along with the appropriate taxonomic category heading or in one of the blank spaces (if the word was from a noncued recall condition). At cued recall, then, subjects were provided with a word from each of the initially processed categories in order to reestablish more of the original learning context. If the presence of category cues during recall in Experiment 1 improved recall because the cues reestablished the learning context, the differences between recall in Conditions C-C and C-U in the present experiment should be minimized.

As in Experiment 1, the categories of words were counterbalanced within conditions. The particular exemplar from each category used to reestablish the learning context was counterbalanced across subjects. As a result of providing subjects with exemplars during cued recall, the maximum number of items to be recalled was one less in each of the various-sized categories than in Experiment 1.

Results and Discussion

The number of items recalled from each of the categories was converted into the percentage of total possible recall. The means are shown in Table 2.

The pattern of results in the four conditions is different from that seen in Experiment 1. Cuing at recall with category headings improved recall only in the C-C condition; recall was poorest in the U-C condition. In addition, any beneficial recall effects from having categorized the items during processing were minimized if the taxonomic headings were not provided at recall. In other words, not only were category headings weak cues at recall if the items were not categorized during processing, but also categorizing the items during process-

Table 2
Mean Percentage of Cued Recall as a Function of Category Size and Cue Presence During Input and Output in Experiments 1 and 2

Input-Output Cuing	Category Size 2		Category Size 3		Category Size 5		Category Size 6	
	Experiment		Experiment		Experiment		Experiment	
	1	2	1	2	1	2	1	2
Cued-Cued	90.90	75.00	86.14	78.12	64.55	60.94	68.64	68.75
Cued-Uncued	61.36	62.50	49.77	59.38	47.27	46.88	52.55	46.25
Uncued-Cued	47.73	31.25	69.36	59.38	59.29	43.75	53.45	47.50
Uncued-Uncued	27.27	68.75	11.86	62.50	33.64	48.44	26.41	37.50

Note—Category sizes in Experiment 2 are one item less than listed due to experimenter-provided exemplar in each condition.

ing had minimal beneficial effects for later recall if the category heading cue was not provided. Cues present during processing which aid in the organization of the material do not substantially improve recall unless those cues are also present to aid in the retrieval of the material.

The lack of a significant effect due to the presence of category headings at recall is not difficult to explain. Since an exemplar from each category was provided at recall, the presence or absence of the category heading would be expected to cause only minimal differences in recall. To the extent that an exemplar resulted in subjects' recalling the taxonomic category, the experimenter-provided category headings were redundant cues. On the other hand, the significant effect upon recall of categorizing the items during processing [$F(1,15) = 4.45$, $p < .05$], as well as the significant interaction of Processing Categorization by Cuing at Recall [$F(1,15) = 7.59$, $p < .05$], indicates that regardless of the redundancy of the recall cues, recall is attenuated if the items are not organized during processing, and particularly so if the organizational cues are also absent during recall.

GENERAL DISCUSSION

The results of the present study do not support a strong statement of the encoding specificity principle. As in the earlier study of Epstein and Dupree (1977), cuing at recall with appropriate category headings improved recall regardless of whether the cues had been present during processing. According to the encoding specificity principle, recall cues should have been ineffective if they were not initially present during processing.

Furthermore, neither a generation-recognition model nor a more general contextual cue model is sufficient to account for the results of this study. The former model predicts equal recall in those conditions cued at recall but not during processing and those cued at both times: recall in the latter condition was significantly better in the present study. The general contextual cue model also cannot account for the generally lower levels of recall when subjects were given additional contextual cues at recall. The increased similarity between the processing and recall contexts should have produced better recall, according to this latter view.

Whereas the original encoding specificity principle is inadequate to account for the results, Tulving's (Note 1) recent modification is more successful. The modification states that recall is best when subjects are cued at recall with a stimulus that was also present during processing; recall may also be improved by presenting subjects with new cues during recall, but not to the same extent as when the cues were also present during processing. Unfortunately, the hypothesis merely describes

a frequently observed occurrence without providing a mechanism to explain why cuing only at recall has any beneficial value at all. That is, the theoretical value of the modified principle is severely limited.

It appears, then, that we do not at present have a general theory that is adequate to deal with the complexities of cued recall. Rather than attempting to formulate a general theory, perhaps greater attention should be devoted to formulation of situationally specific mini-theories to deal with particular encoding tasks and cuing conditions. It is hoped that the present study might stimulate interest in that direction.

REFERENCE NOTE

1. Tulving, E. Personal communication, November 1977.

REFERENCES

- BILODEAU, I. McD., & SCHLOSBERG, H. Similarity in stimulating conditions as a variable in retroactive inhibition. *Journal of Experimental Psychology*, 1951, **41**, 199-204.
- EPSTEIN, M. L., & DUPREE, D. A. Category headings as recall cues: A test of the principle of encoding specificity. *Journal of General Psychology*, 1977, **97**, 101-108.
- HUMPHREYS, M. S., & GALBRAITH, R. C. Forward and backward association in cued recall: Predictions from the encoding specificity principle. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 702-710.
- MARTIN, E. Generation recognition theory and the encoding specificity principle. *Psychological Review*, 1975, **82**, 150-153.
- PELLEGRINO, J. W., & SALZBERG, P. M. Encoding specificity in cued recall and context recognition. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 261-270.
- SANTA, J. L., & LAMWERS, L. L. Encoding specificity: Fact or artifact? *Journal of Verbal Learning and Verbal Behavior*, 1974, **13**, 412-423.
- SANTA, J. L., & LAMWERS, L. L. Where does the confusion lie? Comments on the Wiseman and Tulving paper. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 53-57.
- THOMSON, D. M., & TULVING, E. Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, 1970, **86**, 255-262.
- TULVING, E., & OSLER, S. Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, 1968, **77**, 593-601.
- TULVING, E., & THOMSON, D. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, **80**, 352-373.
- WALLACE, W. P. Recognition failure of recallable words. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, **4**, 441-452.

(Received for publication June 13, 1979.)