*Bulletin of the Psychonomic Society*
*1974, Vol. 4 (1) 9-11*

# The generation and recognition components of encoding specificity*

PHILIP M. SALZBERG† and JAMES W. PELLEGRINO††
*University of Colorado, Boulder, Colorado 80302*

The encoding specificity effect was examined in the Tulving and Thomson (1973) generation-recognition-recall paradigm. The critical comparison involved the relative effectiveness of category cues which were either the same as or different from the input encoding of an item. Additionally, an attempt was made to gain control over any bias in responding to the cues by the manipulation of instructional information about the relevance of each cue. Results showed that the probability of generating target items was slightly superior for same cues. Encoding specificity (defined as the same-different cuing effect) was demonstrated both in the recognition of generated items and in cued recall. The instructional manipulation generally tended to attenuate the same-different effect in recognition, thus, suggesting a response bias component of the effect.

A currently accepted conceptualization of retrieval from memory is a large class of models subsumed under the general heading of the generation-recognition theory (e.g., Bahrick, 1970). Recall in the theory is thought of as two distinct processes: the implicit generation or retrieval of items followed by a recognition decision about each item. A generated item is "recalled" if its mnemonic information exceeds some criterion value, i.e., if it is recognized.

A current issue of some importance for the model concerns the effectiveness of retrieval cues. Bahrick's (1969) account of the facilitative effect of a retrieval cue was essentially that the cue increased the probability of generating the target item. The process of recognition presumably was not affected by the cue. In contrast to this view, the encoding specificity principle and some data presented in its support (Tulving & Thomson, 1973) would localize the effect of a cue in the recognition stage. Tulving and Thomson had Ss first study a word list where each item was paired with a weak associative cue. Ss were then given strong associates of. the target items and asked to generate associative responses. Next, they were asked to examine their generated words and circle any they thought were target items. Finally, they were given the original weak associates as retrieval cues and asked to recall the targets. Thus, their testing sequence was generation-recognition-recall. The results showed no effect of the input procedure on the probability of generating the items; target items given as primary responses to the strong cues matched their frequencies in associative norms. However, the probability of recognizing the generated target items was only 20%-30% as compared to cued recall performance of about 60%. From these and other data (Thomson & Tulving, 1970), it was argued that the effectiveness of a retrieval cue is determined by the extent to which the target was encoded with respect to that cue at input; and further, that the locus of cue effectiveness is in recognition, not in generation as proposed by Bahrick (1969).

It would seem, however, that the latter conclusion could be justified only if generation performance could be shown not to differ for cues which were and were not presented at input. Under the Tulving and Thomson (1973) generation-recognition-recall procedure, comparisons between performance to weak and strong (old and new) cues are necessarily confounded with different stages of practice and different tasks. Moreover, their generation task does not involve any of the original input cues. Thus, one purpose of the present experiment was to provide a direct comparison of old with new cues by modifying the Tulving and Thomson (1973) procedure so that half of the present cues were old cues (category labels representing the same meanings of the items as the input encoding), while the other half were new cues (category labels representing different meanings than those encoded at input). Thus, the present procedure permitted comparisons between old vs new cues for both the generation and recognition subprocesses of recall.

Another factor in the Tulving and Thomson paradigm that must be considered concerns differences between their words representing weak and strong associative relationships. While several different types of associations exist in the Tulving and Thomson lists, one clear difference between their weak and strong associative relationships is that the former never involve antonyms while almost 50% of the latter are of this

*This is publication No. 46 of the Institute for the Study of Intellectual Behavior, University of Colorado, and is based on research supported by Grants GB-37077x and GB-25433 from the National Science Foundation, as well as by the second author's U.S.P.H.S. fellowship, No. MS-8030-015, and by Grant MG-19577 from The National Institute of Mental Health. The authors would like to thank William F. Battig, who sponsors this paper and takes full editorial responsibility for its contents, and Gregory A. Kimble for their support and encouragement of this research.

†Reprint requests should be sent to Philip M. Salzberg, Department of Psychology, University of Colorado, Boulder, Colorado 80302.
††Now at the University of Pittsburgh.

**Table 1**
**Percent Correct Recall for Same and Differently Cued Items as a Function of Test Sequence**

| Test Sequence | Item Cuing | | Mean |
|---|---|---|---|
| | Same | Different | |
| CR First | 73.8 | 43.4 | 58.6 |
| CR Last | 60.4 | 41.7 | 51.0 |
| FR First | 35.4 | 44.3 | 39.8 |
| Mean | 56.5 | 43.1 | |

*Note–Same and different item cuing is a pseudovariable in the free recall test.*

type. Part of the depressed performance during their recognition task may be a function of differential responding based upon general contextual information in the original input list. Thus, the present procedure eliminated this confounding by using homograph target items which were high-frequency members of one or more taxonomic categories. Rather than weak and strong associative cuing where the cue-target relationship is variable, the cue-target relationship both at input and output was always category to instance in the present lists. This avoids the possibility of differential responding based upon general semantic contextual list features. It does not, however, prevent the possibility of differential criteria in responding to category cues that did and did not occur in the input list. It is reasonable to expect that Ss might be more cautious in responding to cues they do not recognize as part of the input list. To examine the possibility of such a bias in recognition, two levels of information about the relevance of each cue were included. Finally, the present design also examined different testing sequences to determine if they would differentially affect cued recall and recognition of generated items performance with respect to the magnitude of the old-new cuing effect.

## METHOD

### Design

The design involved six independent groups in a 3 x 2 x 2 mixed factorial. One of the between-Ss factors consisted of three testing sequences. These were the Tulving and Thomson generation-recognition-cued recall sequence, a cued recall-generation-recognition sequence, and a free recall-generation-recognition sequence. These sequences will be designated as CR last, CR first, and FR first. The other between-Ss factor was complete vs incomplete information about the cue-target relationship during the recognition test. The within-Ss factor was type of item cue, which referred to either the same or a different meaning of the target item from that encoded at input for equal halves of the list.

### Materials

The to-be-remembered items were 32 homographs, each with two sematically distinct meanings. The category cuing procedure required that all items be members of two distinct categories. However, extensive efforts to obtain items meeting this requirement did not yeild a sufficient number of words which had a strong relationship to both categories. Thus, it was impossible to counterbalance for items and meanings across the same vs different cuing conditions. As a result, the list consisted of one set of 16 items always cued by same cues and another set always cued by different cues. However, the two item sets were comparable in terms of the preexperimental degree of relatedness between the cues and target items. For those items included in the Battig and Montague (1969) norms, the mean rank order and taxonomic frequency of the items within their output categories was 7.4 and 186.8 for the different set and 8.5 and 165.1 for the same set. Those items not in the norms (6 from each set) also showed similar characteristics in normative data provided by 20 Ss.

### Procedure

Ss were presented slides containing incomplete sentences of the form "x and Y are" (e.g., "robin and CARDINAL are") and were required to write down a category label that expressed the relationship between the two words. In addition, Ss were told to remember the Y item which was in capital letters. The items were presented for 12 sec each, during which S gave his category response and studied the capitalized memory word. Whereas Thomson and Tulving (1970) used "priming" lists in order to induce S to encode the target with respect to the cue, the present procedure contains the encoding inducement in the category response.

For the cued-recall test, sheets of paper were provided which contained 32 category labels (16 same and 16 different). Recall instructions emphasized that each category was related to one of the originally presented capitalized words and that S should use these cues as aids to recalling the target words. Free-recall Ss were told only to write down as many capitalized words as they could remember on a sheet containing 32 lines. Three minutes were allowed for the cued and free recall tests.

In the item-generation task, Ss were given test booklets containing the 16 same and 16 different category labels with 12 lines for each category. The instructions emphasized that this task was designed to examine which words people generally include as members of various categories, and that this task was unrelated to the previous tasks. Ss were given 24 min to write down as many instances of each category as they could. In the recognition test immediately following generation, all Ss were told to examine the words they had given and to circle any that they recognized as old words. Also, they were told that the category names might help them to do this. The Ss in the complete-information condition were told in addition that each category was related to one of the originally capitalized words, so that they should examine each category carefully for possible old words.

### Subjects

There were 18 Ss in each of the six experimental groups. A group testing procedure was used with 5-18 Ss in a particular condition tested together. All Ss were University of Colorado undergraduates participating to fulfill an introductory psychology class requirement.

## RESULTS

### Recall

The important consideration in recall was the magnitude of the cuing effect for same and differently cued items as compared to free-recall performance. As can be seen in Table 1, there was an interaction between item cuing and test sequence, $F(2,102) = 41.05$, $p < .001$, which is entirely due to the facilitative effect of same cues in both cued recall groups and the absence of any effect of different cues. The facilitation for same cued items was also sufficient to produce a significant main effect of cues, $F(1,102) = 54.80$, $p < .001$. Also significant was the test sequence effect as shown in the last column of Table 1, with CR-first better and FR-first worse than the Tulving-Thomson CR-last condition, $F(2,102) = 13.12$, $p < .001$.

### Generation and Recognition

The analysis of proportion of target items generated showed that this was closely related to recall, being highest for the CR-first groups (79.4%) than either the CR-last (74.1%) or FR-first groups (71.5%), $F(2,102) = 7.08$, $p < .01$. Also, the proportion of target items generated to same (77.6%) was greater than to different (72.4%) category cues, $F(1,102) = 10.44$, $p < .01$.

Recognition performance was conditionalized on target occurrence in generation, and the percentages of correct recognition for same and differently cued items are shown in Table 2. All main effects and interactions were significant. The CR-first test sequence shows significantly higher overall performance than the other two sequences, $F(2,102) = 8.11$, $p < .01$. Complete instructions are significantly above incomplete instructions, $F(1,102) = 4.34$, $p < .05$, and the interaction of instructions with test sequence, $F(2,102) = 3.40$, $p < .05$, confirms that complete information only improved performance in the CR-last and FR-first sequences. This outcome is expected since the CR-first groups were given complete information in the cued recall test prior to the recognition task.

Of greater importance is the marked superiority of same over different item cuing, $F(1,102) = 74.19$, $p < .001$, which showed second- and third-order interactions with the between-Ss variables. These interactions (ps $< .05$) indicate that the difference between same and different cues varies with instructions such that it is generally reduced under complete information. However, the magnitude of this reduction varies across test sequences such that there is a reversal in the CR-first sequence, a small reduction in the CR-last sequence, and a sizable reduction in the FR-first sequence.

**Table 2**
**Percent Correct Recognition of Generated Items for the Three Test Sequences, Complete vs Incomplete Recognition Information, and Same vs Different Cues**

| Test Sequence | Recognition Information | Item Cuing: Same | Item Cuing: Different | Mean |
|---|---|---|---|---|
| CR First | Complete | 82.5 | 67.0 | 74.7 |
| | Incomplete | 86.2 | 74.0 | 80.1 |
| CR Last | Complete | 79.4 | 55.5 | 67.4 |
| | Incomplete | 71.0 | 41.1 | 56.0 |
| FR First | Complete | 72.3 | 70.2 | 71.2 |
| | Incomplete | 70.7 | 40.2 | 55.4 |
| Mean | | 77.0 | 58.0 | |

## DISCUSSION

The present results clearly replicate the Tulving and Thomson (1973) generation-recognition-recall results and extend them to taxomonic cuing conditions. The encoding-specificity effect in both recall and recognition seems also to be unaffected by order of testing or by the simultaneous occurrence of old and new cues. The test sequence effect that did appear involved an overall increase in performance which may simply reflect the shorter time period between input and initial test in the CR-first groups as compared to the CR-last groups. Thus, whether the task is cued recall or item recognition following generation, there is a sizable effect of type of cuing. An encoding specificity effect in cued recall tasks would therefore appear to result mainly from the effect of the cue in the recognition rather than generation stage of recall. Caution must be observed in making such a strong statement since there apparently is a cumulative effect of both stages. The present data show a small difference in the generation probabilities to old and new cues. However, the largest performance effect comes from the failure to recognize an item if it is contained within a new associative or semantic context (see also Light & Carter-Sobell, 1970) even though S is responsible for producing the item.

It is reasonable to ask whether the effect is a genuine discriminability problem in recognition or simply a bias in responding to familiar and unfamiliar cues. The present manipulation of test information in recognition tended to show that such a bias does operate in recognition, although it is difficult to gain control over this bias simply by instructions to attend to each category set. Further evidence in support of the operation of a bias is the difference in false alarms to same (22.5%) and different category sets (9.5%). If such a bias operates in this type of recognition task, then the effect may be attributable to criterion differences in responding to old vs new cues, rather than to any real difference in discriminability or sensitivity. Also, if the bias is assumed to operate in cued recall as well as in recognition, then the old-new cuing effect may result simply from S's lack of responding or unwillingness to respond to a new cue, i.e., a new cue may be totally ignored. Thus, the present data suggest three possible sources of the encoding specificity effect in cued recall: an initial cue recognition stage, the generation stage, and the item recognition stage. Although it is unclear from the present data whether the effect in the last stage is due to a response bias or a problem in recognition sensitivity, other work (Pelligrino & Salzberg, 1974) has shown both components to be operative. Thus, the encoding specificity effect in recognition seems to be attributable both to response bias and to some real differences in mnemonic properties of old and new cues.

## REFERENCES

Bahrick, H. P. Measurement of memory by prompted recall. Journal of Experimental Psychology, 1969, 79, 213-219.
Bahrick, H. P. Two-phase model for prompted recall. Psychological Review, 1970, 77, 215-222.
Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph, 1969, 80, No. 3, Part 2.
Light, L. L., & Carter-Sobell, L. Effects of changed semantic context on recognition memory. Journal of Verbal Learning & Verbal Behavior, 1970, 9, 1-11.
Pellegrino, J. W., & Salzberg, P. M. Encoding specificity in cued recall and context recognition. Journal of Experimental Psychology: Human Learning & Memory, 1974, submitted.
Thomson, D. M., & Tulving, E. Associative encoding and retrieval: Weak and strong cues. Journal of Experimental Psychology, 1970, 86, 255-262.
Tulving, E., & Thomson, D. M. Encoding specificity and retrieval processes in episodic memory. Psychological Review, 1973, 80, 352-373.

(Received for publication April 15, 1974.)