

Depth of processing: When the principle of congruity fails

DANIEL L. ROENKER

Western Kentucky University, Bowling Green, Kentucky 42101

STEVEN K. WENGER and CHARLES P. THOMPSON

Kansas State University, Manhattan, Kansas 66502

and

BRIAN WATKINS

Millikin University, Decatur, Illinois 62522

Two factors, level of processing (Craik & Lockhart, 1972) and congruity (Schulman, 1974), known to have large effects on the recall of experimenter-provided responses to questions, were examined in a series of five incidental learning experiments using subject-generated responses. The data show that manipulation of level of processing has the same effect on recall of subject-generated responses as it does on experimenter-provided responses. However, the effect of congruity is reversed for subject-generated responses. The data suggest that the difficulty of generating unrelated responses ("incongruous" items) may account, at least in part, for the failure of the "principle of congruity" with subject-generated responses.

Since the appearance of articles by Cermak (1972) and Craik and Lockhart (1972), considerable attention has been given to the "depth-of-processing" approach to memory research. Within this model information processing is viewed as a continuum ranging from perceptual analysis, which primarily deals with structural or physical features, through semantic and cognitive analyses. The major tenet of the theory is that deeper (i.e., semantic) or more elaborate processing leads to a more stable memory trace and hence facilitates retention. Evidence supporting this type of model has been rapidly accumulating (e.g., Craik & Tulving, 1975; Epstein, Phillips, & Johnson, 1975).

The most common experimental technique used to explore this model is the incidental learning paradigm. In a typical procedure, the subject is presented with a question stem and a response. The subject's task is to judge whether or not the provided response is an appropriate answer to the question. The variable of interest in such studies is the type of cognitive processing involved in deciding the appropriateness of the target item. Type of processing is presumably manipulated by the nature of the question stem (i.e., structural, phonemic, or semantic). Numerous investigations (e.g., Craik & Tulving, 1975; Hyde & Jenkins, 1973; Till & Jenkins, 1973) have clearly demonstrated that manipulations of

the type of question stem have marked consequences for subsequent recall of the target items.

To date, depth-of-processing studies have dealt with to-be-remembered target items provided by the experimenter. The present research was designed to extend these effects to a situation in which the subject generated the to-be-remembered target items. The extension is nontrivial because the depth-of-processing hypothesis demands that the effect hold whether the items are produced by the subject or by the experimenter. That is, processing at the level requested (e.g., phonemic) must occur whether the subject is asked to generate an appropriate item or to determine whether a given item is appropriate. It is possible, however, that all self-produced items will be equally easy to recall.

The present procedure is also of interest because it seems likely that subject-generated responses will produce an exception to what has been dubbed the "principle of congruity" (Craik & Tulving, 1975; Schulman, 1974). That is, with experimenter-provided responses, items that belong to the set described by the accompanying question (congruous items) are recalled better than items that are not set members (incongruous items). Craik and Tulving (1975) suggest that the query provides an encoding context with which a congruent target word may be easily integrated. Then, during recall, the integrated unit may be easily reinstated by representing either part of the unit. Following that reasoning, it is clear that the question stem *Is a tree?* with the response "oak" would be regarded as a well-integrated unit, with "oak" as a congruent response. On the other hand, the question stem *Is not a tree?* with

The authors wish to thank David Freed for his assistance in the data collection phase of Experiment 1. Requests for reprints should be sent to Daniel L. Roenker, Department of Psychology, Western Kentucky University, Bowling Green, Kentucky 42101.

the response "tiger" would probably not be a well-integrated unit, and thus "tiger" should be classified as an "incongruent" response. Given that interpretation, there are at least two reasons why subject-generated responses might produce better recall for incongruent than for congruent responses.

First, the question stem probably initially directs the search for a member of the set designated by the positive form of the question. That is, given that questions ordinarily are designed to elicit an answer from a positively defined subset of items, both *Is a tree?* and *Is not a tree?* will probably tend to direct memory search toward the subset of tree names. It would, therefore, be somewhat more difficult for a subject to generate a nonmember of the set defined by the (positive) question stem than to generate a member of that set. Such additional cognitive effort could well lead to superior retention in subsequent free recall (Gardiner, Craik, & Bleasdale, 1973).

Second, in the case of negative questions, subjects may adopt the strategy of selecting responses from a pool of highly familiar and interrelated items. Because responses to positive questions will not be interrelated, one might expect better recall for the "nonmember" items.

The initial experiment was conducted to replicate the depth-of-processing and congruity effects with experimenter-provided responses and to determine whether the same effects would obtain with subject-generated responses.

EXPERIMENT 1

Method

Subjects. The subjects were 20 undergraduate students enrolled in a course in human learning and memory at Millikin University. All subjects were tested as a group.

Design. The experimental design had 12 treatment conditions, all manipulated within subjects. Six of the treatments represented a 3 by 2 factorial design, crossing question type (i.e., structural, rhyming, or categorical) with response type (i.e.,

member or nonmember of the set defined by the question). In these six conditions the subject was given a question stem and a target item. The subject's task was to judge whether or not the target item was a member of the set defined by the question. These conditions are approximately equivalent to those used by Craik and Tulving (1975).

The remaining six conditions represented a 3 by 2 factorial design, crossing question type with response type (i.e., member, nonmember of the set defined by the positive form of the question). The 12 conditions and sample questions are illustrated in Table 1. It should be noted that in the latter six conditions the subject was required to produce an item in response to the question stem.

Materials. A total of 60 question stems were constructed so that they could be equally divided into three classes: structural, rhyming, or categorical. For each of the 60 question stems, four variations were constructed: (1) the question stem and a response representing a member of the set defined by the question, (2) the question stem and a nonmember response, (3) the question stem worded to request a member of the set defined by the question, and (4) the question stem worded to request a nonmember of the set defined by the positive form of the question. For a structural question the four variations may be seen in lines 1, 2, 7, and 8 of Table 1. Each variation of each question was counterbalanced across sets of four subjects such that each question stem appeared in a different form for each of the four subjects. Each version of each stem appeared equally often across subjects, and combinations of basic question stems were randomized across sets of four subjects. All experimenter-provided target items were equated for Thorndike-Lorge (1944) frequency.

Procedure. Each subject was given a deck of 60 index cards (7.6 x 12.6 cm). Each deck contained five instances of each of the 12 conditions arranged in a random order. The top question was covered by a blank card. The subjects were then instructed that the task dealt with their ability to recognize information already stored in their memory and with their ability to produce information upon demand. They were not told that they would later be given a free recall test. Subjects were informed of the two basic types of questions contained in the deck. Some cards contained a question with an answer typed beneath in capital letters. The subjects were to judge the appropriateness of the answer by circling either Y or N (for yes or no, respectively) in the lower right-hand corner of the card. The remaining cards merely contained a question stem and a blank line located below the question. The subjects were to respond to the question by writing a response in the space provided. The subjects progressed through the deck at the rate of one card every 9 sec. Previous cards were turned face down in a separate stack. After the last

Table 1
Sample Questions and Responses as a Function of Conditions

Condition		Question	Answer
Experimenter-Provided Responses			
Structure	Yes	Contains P and K?	Pancake
Structure	No	Contains P and K?	Wrinkle
Rhyme	Yes	Rhymes with STRIFE?	Wife
Rhyme	No	Rhymes with STRIFE?	Career
Category	Yes	An INSECT?	Roach
Category	No	An INSECT?	Circuit
Subject-Generated Responses			
Structure	Member	Contains P and K?	
Structure	Nonmember	Does not contain P and K?	
Rhyme	Member	Rhymes with STRIFE?	
Rhyme	Nonmember	Does not rhyme with STRIFE?	
Category	Member	An INSECT?	
Category	Nonmember	Not an INSECT?	

card, the decks were collected and labeled by subject number. Following collection of the cards, answer sheets were distributed and subjects informed of the free recall test. Approximately 3 min elapsed between answering the last question and free recall. The subjects were given 10 min in which to complete their free recall.

Results

The number of correctly recalled words during free recall served as the basis for analysis. Data from conditions in which the subject generated the response were analyzed separately from conditions in which the experimenter provided the to-be-remembered item. Further, since the experimental design permitted the subject to produce the same response to more than one question, two methods of scoring were employed. In both methods, each subject's score for each condition was computed by dividing the number of words correctly recalled by the number of responses originally produced. Subjects failed to produce a correct response 3.9% of the time, and these incorrect responses were excluded from the scoring. Under the liberal scoring method, all responses recalled in free recall which had been originally produced in response to more than one question, or which overlapped with experimenter-provided responses, were scored as correct for all appropriate questions regardless of question type. Under the conservative scoring method, all duplicate responses were excluded from the analysis. Duplicate responses accounted for approximately 11% of the free recall data. Since the results were equivalent for both the liberal and conservative analyses, only the results based on the conservative method will be presented.

Experimenter-provided responses. The mean probability of recall as a function of question type and response type are shown in the top portion of Table 2 for experimenter-provided to-be-remembered items. An analysis of variance performed on these data revealed only a significant main effect of question type [$F(2,38) = 10.92, p < .001, MSe = .014$], indicating that the amount recalled increased with level of processing. This finding is consistent with the data reported by Craik and Tulving (1975). Although the effect of response type (member-nonmember) did not quite reach significance [$F(1,19) = 3.64, .10 > p > .05, MSe = .012$], the means are in the same direction as those reported by Craik and Tulving (1975).

Subject-generated responses. The mean probability of recall as a function of question type and response type is shown in the lower portion of Table 2 for subject-generated items. The effect of response type was significant [$F(1,19) = 19.62, p < .001, MSe = .053$], with nonmembers being recalled at a higher rate than members. This finding represents a reversal of the response type (member-nonmember) effect typically found for experimenter-provided items. In addition, the effect of question type approached, but did not quite reach, significance [$F(2,38) = 2.86, .10 > p > .05, MSe = .059$].

Table 2
Mean Final Free Recall Performance as a Function of Question Type and Response Type for Experimenter- and Subject-Generated Responses

Question Type	Response Type		
	Member	Nonmember	Mean
Experimenter-Provided Responses			
Structure	.07	.05	.06
Rhyme	.11	.11	.11
Category	.23	.13	.18
Mean	.14	.10	
Subject-Generated Responses			
Structure	.42	.53	.48
Rhyme	.34	.58	.46
Category	.48	.69	.59
Mean	.41	.60	

Finally, an examination of Table 2 shows that recall was considerably higher for subject-generated items. This most likely reflects the fact that subject-generated items are more likely to be idiosyncratically easier items than those provided by the experimenter.

Discussion

The experiment produced two results of interest. First, the well-established advantage for member (congruous) target items in experimenter-provided conditions was replicated, although it was only marginally significant in this study. However, that advantage was clearly reversed in the subject-generated condition. Thus, it appears that the "law of congruity" fails for subject-generated responses.

Second, the effect of level of processing was in the expected direction in both conditions but was only marginally significant in the subject-generated condition. The question thus remains as to whether the levels-of-processing manipulation is effective in an incidental learning task in which the target items are subject generated.

While these data are of interest, the rather high rate of response duplication (11%) in the subject-generated condition is bothersome. Thus, it seemed prudent to replicate the subject-generated portion of the experiment with specific instructions to the subjects to avoid duplication of responses. It should also be noted that the replication was conducted in a different laboratory with independently generated materials.

EXPERIMENT 2

Method

Subjects. The subjects were 24 students enrolled in introductory psychology at Kansas State University. All were given extra credit for participating. They were tested in two groups of approximately 12 members each.

Design. A within-subjects design was used, with six conditions resulting from the factorial combination of three levels of question type (i.e., structure, rhyme, or category) with two levels of response type (i.e., member or nonmember requested).

Materials. Ten items were constructed for each of the three

types of questions. The 10-item sets required analysis of structure (e.g., "Contains L & N?"), rhyme construction (e.g., "Rhymes with bed?"), and category analysis (e.g., "Is a flower?"), respectively. All questions were experimenter generated with the categories chosen from the Battig and Montague (1969) norms. Each subject received a booklet of 30 cards (10.2 x 12.7 cm) with a single question along the top of each card and the words "YES" and "NO" on the bottom of each card. For each question type, one half of the cards had "YES" circled and the other half had "NO" circled. Each subject's 30 cards, then, consisted of five cards from each of the six possible combinations of question type and response (i.e., YES-NO) type. These cards were randomly ordered, with the restriction that no more than two consecutive cards be of the same question type or response type. Two forms of the booklet were constructed so that, over all booklets, the responses YES and NO appeared equally often with each question.

Procedure. The subjects were instructed that they were participating in a pilot study to determine the "natural" way in which people respond to particular questions or words. Subjects were not told that they would later be asked to recall the words they were to generate. The task of the subject was to respond to each combination of question and response with an appropriate word. For example, if the question was "Is a flower?" and YES was circled, the subject was to respond with the name of the flower. If NO was circled with the same question, the subject was to respond with any word that was not the name of a flower. Subjects were instructed to use a different word to answer each question. After giving a sample question for each question type, the task was begun. Subjects were given 9 sec to respond to each question. The response was written on the card following the question. At the end of the 30 questions, the booklets were collected and blank sheets for recall were distributed. This procedure took about 3 min. The subjects were then asked to recall all of the generated words they could remember. These were written on the recall sheet. Ten minutes were given for this final recall.

Results

Instructions to the subject (together with the elimination of the experimenter-provided responses) had the desired effect in that duplicate responses accounted for less than 1% of the free recall data. The analysis of the data excluded any incorrect responses (less than 1%) and the duplicate responses. Again, each subject's score for each condition was computed by dividing the number of words correctly recalled by the number of responses originally produced. Table 3 shows the mean probability of recall as a function of question type and response type. As before, the effect of response type was significant [$F(1,23) = 6.33, p < .05, MS_e = .055$], with nonmembers being recalled at a higher rate than members. In addition, there was a reliable effect of question type [$F(2,46) = 6.72, p < .01, MS_e = .063$]. Subsequent Newman-Keuls analyses established that responses for category questions were recalled reliably better than responses for structure or rhyme questions, which did not differ reliably. There was no reliable interaction of Question Type by Response Type.

Discussion

The results replicate Experiment 1 data for subject-generated responses with the minor exception that the present data show a statistical, as well as numerical,

Table 3
Mean Probability of Recall as a Function of Question Type and Response Type for Experiment 2

Question Type	Response Type		
	Member	Nonmember	Mean
Structure	.39	.42	.41
Rhyme	.28	.43	.36
Category	.48	.59	.53
Mean	.38	.48	

effect of levels of processing. Thus, the well-established and substantial effect of processing level can be demonstrated in an incidental learning task using subject-generated responses.

At this point it also seems clear that the superiority in recall for member (congruous) items obtained with experimenter-provided items is reversed when dealing with subject-generated responses. This reversal appears to be relatively robust in that it can be obtained under various levels of processing.

It is possible, however, that we are completely incorrect in our assumption that nonmember items are not well integrated with the question stem and, hence, should be classified as incongruent responses. Subjects could select response items that are not members of the category defined by the question but which are highly related nonetheless. For example, the response to *Not a tree?* might be "leaf." Such nonmember responses could conceivably be better integrated with the question stem than member (i.e., category) responses and, for the reasons suggested by Craik and Tulving (1975), produce better recall. That possibility is readily tested by comparing cued and free recall. If nonmember responses are somehow better integrated with the question stem than are member responses, the advantage of nonmember over member responses should increase when the question stem is provided as a cue at the time of recall.

EXPERIMENT 3

Method

Subjects. The subjects were 48 introductory psychology students at Kansas State University who participated to fulfill a course requirement. Subjects were tested in groups of approximately six members.

Design. A one-between, two-within factorial design was used. The between-subjects factor was type of recall (free or cued). The within-subjects factors were question type (category or rhyme) and response type (member or nonmember). The combination of the three factors resulted in eight conditions.

Materials. Twenty question stems were constructed for each of the two types of questions. The materials were generated independently except that the categories were again chosen from the Battig and Montague (1969) norms. Within each question type, each question was written both in the form requesting a member response (e.g., "A sport?" or "Rhymes with ball?") and in the form requesting a nonmember response (e.g., "Not a sport?" or "Does not rhyme with ball?").

Each subject received a booklet of 40 cards (10.2 x 12.7 cm),

with a single question appearing on each card. Each booklet consisted of 10 cards from each of the four possible combinations of question type and response type. The cards were randomly ordered, with the restriction that no more than two consecutive cards be from the same condition. Each category and rhyme question stem appeared once and only once in each booklet. Four forms of the booklet were constructed so that, over all booklets, the member and nonmember forms of the question stem appeared equally often.

Procedure. The procedure was identical to that in Experiment 2, with one necessary exception for subjects participating in the cued recall group. During recall, these subjects were given an answer sheet containing all of the question stems (randomly ordered) and asked to write down for each question the same response they had given previously. They were further instructed to write down at the bottom of the sheet any items they recalled but could not pair with the appropriate question stem. Both the cued and free recall groups were given 10 min for the recall task.

Results

Because the hypothesis being tested demands that the subject pair the recalled item with the appropriate question stem, three errors of commission could occur in recall. Specifically, items could be incorrectly paired with a question stem (incorrect pairs), items could be recalled that were not previously generated (intrusions), and items could be recalled separately (unpaired items). Incorrect pairs, intrusions, and unpaired items accounted for 5.7%, 9.1%, and 3.8%, respectively, of all recalled items. All three types of errors were excluded from the data analysis. Duplicate responses (2.2%) and incorrect responses (less than 1%) were also excluded from the analysis.

The mean probability of recall as a function of conditions is shown in Table 4. Each subject's score for each condition was computed by dividing the number of words correctly recalled in that condition by the number of responses originally produced for that condition. An analysis of variance performed on these data indicated a highly significant Response Type by Type of Recall interaction [$F(1,46) = 129.00$, $p < .001$, $MSe = .032$]. Tests for simple main effects of response type indicated that recall for nonmembers was significantly higher than recall for members in the free recall condition [$F(1,23) = 12.43$, $p < .01$, $MSe = .022$], while recall for members was significantly higher than recall for nonmembers in the cued recall condition [$F(1,23) = 132.22$, $p < .001$, $MSe = .041$]. Further, a separate analysis performed on nonmember responses only showed that performance reliably decreased under cued as compared to free recall [$F(1,46) = 4.13$, $p < .05$, $MSe = .060$]. These findings are incompatible with the hypothesis that subjects select nonmember responses that can be better integrated with the question stem than member responses.

As expected, there was a significant main effect of question type, with category questions recalled better than rhyme questions [$F(1,46) = 33.67$, $p < .001$, $MSe = .028$], and the overall level of recall performance was significantly higher in the cued recall condition than

Table 4
Mean Probability of Recall as a Function
of Conditions for Experiment 3

Question Type	Response Type		
	Member	Nonmember	Mean
	Free Recall		
Rhyme	.37	.46	.42
Category	.46	.57	.51
Mean	.41	.52	
	Cued Recall		
Rhyme	.84	.29	.57
Category	.95	.54	.75
Mean	.90	.42	

in the free recall condition [$(1,46) = 27.65$, $p < .001$, $MSe = .064$]. Other than the main effect of response type, which becomes irrelevant in view of the Response Type by Type of Recall interaction noted above, there were no other statistically significant main effects or interactions.

Discussion

These data clearly show that nonmember recall is not improved by cuing. Thus, they are inconsistent with the hypothesis that, for nonmember questions, subjects are choosing items uniquely related to the question stem or some portion of it.

Given the demands of the task in this experiment, it is not surprising that nonmember recall reliably decreased (instead of being enhanced) under cued recall. The task requirements are such that most subjects probably generate nonmember items that are unrelated to the question stem. As a result, the question stem will be a poor retrieval cue, and it may be difficult during cued recall for subjects to pair available nonmember items with the appropriate question stems. We might note that, although our instructions urged subjects to list separately items that could not be paired, few items were so listed. Whether that reflects the inaccessibility of additional items or a criterion decision by the subject remains to be determined. In either case, the present data clearly demonstrate that providing cues during recall does not facilitate recall of nonmember items.

As noted in the introduction, one clear possibility that might account for member-nonmember differences in recall is that the subject may choose nonmember items from some idiosyncratically related pool of items. Whenever the subject is faced with choosing a nonmember item, he or she may select an item unrelated to the question stem, but related to other nonmember items already produced in response to other questions. For example, the subject may generate all or most nonmember items from a pool of responses representing things in the room, or types of automobiles, etc. Experiment 4 was designed to evaluate this possibility by looking for organizational patterns in recall.

EXPERIMENT 4

Method

Subjects. The subjects were 30 introductory psychology students at Western Kentucky University who participated for extra course credit. All subjects were tested as a group.

Design. A within-subjects factorial design combining question type (rhyme or category) with response type (member or nonmember) was employed.

Materials. The rhyme and category question stems from the subject-generated response conditions of Experiment 1 were used in this experiment.

Procedure. The procedure was essentially the same as that in Experiment 2, with one exception. Instead of standard free recall instructions, subjects were given modified free recall instructions patterned after Buschke (1977). Subjects were provided with an answer sheet lined off into five columns, and were instructed to recall the words in any order they wished, going down the first column of the answer sheet. However, if several words seemed to occur to them together, or if words were recalled as a group, these words were to be written across the same line of the answer sheet. This procedure permitted the assessment of the extent to which items were seen as related by the subject.

Results

Recall. Once again, incorrect responses (4.8%) and duplicate responses (3.0%) were excluded from the analyses. The recall data were analyzed as before to assure that the major effects were replicated in the present study. As can be seen from the upper portion of Table 5, there was an effect of question type, with category responses being recalled at a higher rate than rhyme responses [$F(1,29) = 15.06, p < .01, MSe = .017$]. As expected, nonmembers were recalled better than members [$F(1,29) = 40.31, p < .001, MSe = .015$]. The Question Type by Response Type interaction was also significant [$F(1,29) = 13.09, p < .01, MSe = .016$]. This presumably reflects the fact that the member-nonmember difference was smaller for category questions than for rhyme questions.

Organization. The data were also scored to determine the probability of an item being recalled as a member of a cluster. That is, for each condition (e.g., rhyme nonmember) for each subject, the number of clustered items for that condition (items grouped with at least one other item) were divided by the total number of items recalled for that condition. These data are presented in the lower half of Table 5. If subjects are selecting nonmember items from some idiosyncratically homogeneous set, then that strategy should be reflected in a higher probability of these items appearing in a cluster. As is apparent from Table 5, nonmembers were no more likely to be grouped in recall than members [$F(1,29) < 1, MSe = .093$]. Further, the analysis showed neither a question type effect [$F(1,29) = 1.62, p > .10, MSe = .048$] nor a Question Type by Response Type interaction [$F(1,29) < 1, MSe = .071$].

Discussion

The recall data from Experiment 4 replicate the

Table 5
Mean Probability of Recall and Mean Probability of
an Item Appearing in a Cluster as a Function of
Conditions for Experiment 4

Question Type	Response Type		
	Member	Nonmember	Mean
	Recall		
Rhyme	.26	.48	.37
Category	.43	.49	.46
Mean	.35	.49	
	Clustering		
Rhyme	.16	.24	.20
Category	.25	.25	.25
Mean	.21	.25	

results of the previous three experiments. However, there are no differences among the conditions in clustering behavior in recall. This pattern of results strongly suggests that subjects are not selecting items from some homogeneous pool of related items when required to produce a nonmember response.

There remains but one of the possibilities mentioned in the introduction that might account, at least in part, for the superior recall of nonmember responses. That is, the differences in recall may be due to the amount of cognitive effort involved in the selection and generation of responses. If the selection of nonmembers is more difficult than selection of member responses, then it seems reasonable to expect that response latencies should reflect this difference. That is, it ought to take more time to select and generate a nonmember than a member. Experiment 5 was designed to test that possibility.

EXPERIMENT 5

Method

Subjects. The subjects were 24 general psychology students at Kansas State University who participated to fulfill a course requirement.

Design. A within-subjects design was used, having four conditions resulting from the factorial combination of question type (i.e., category or rhyme) and response type (i.e., member or nonmember).

Materials. The materials were identical to those used in Experiment 3 with the following exception. The questions were recorded by a male speaker on a tape recorder, with a 9-sec interval between the onsets of successive questions.

Procedure. Each subject was tested individually. The questions were presented by tape recorder and the subjects were instructed to respond aloud as quickly and as accurately as they could with a word that correctly answered the question. They were instructed to use a different word to answer each question. The subject's responses along with the questions were recorded and the reaction time between the end of each question and the onset of each response was measured. Each subject's reaction time to each of the 40 questions was measured three times. The mean of these three measures was used as the best estimate of the reaction time for each question. Reaction times for incorrect responses, repeated responses, and failures to respond were dropped from the analysis.

Results

Failures to respond, incorrect responses, and repeated responses accounted for 1.5%, 1.2%, and 2.0%, respectively, of all responses. The average reaction time (RT) for each correctly answered question was computed and the mean RT for each condition was calculated for each subject. The mean RT, pooled across subjects, as a function of conditions is shown in Table 6. As can readily be seen, nonmember responses took considerably longer to produce than did member responses [$F(1,23) = 89.01$, $p < .001$, $MSe = .151$]. On the other hand, the mean RT for category questions was not reliably different from that for rhyme questions [$F(1,23) = 2.45$, $p > .05$, $MSe = .101$]. Finally, the Question Type by Response Type interaction was significant [$F(1,23) = 12.63$, $p < .01$, $MSe = .083$]. A simple main effects analysis revealed that the member-nonmember differences were significant in both the category and rhyme conditions [$F(1,23) = 133.04$, $p < .001$, $MSe = .083$, and $F(1,23) = 22.92$, $p < .001$, $MSe = .152$, respectively]. Thus, the interaction reflects the fact that member-nonmember differences in RT were larger for category questions than for rhyme questions. It is clear that the RT data are consistent with the notion that selection and generation of nonmembers is more difficult than selection and generation of members.

It should be mentioned that, as Craik and Tulving (1975) have demonstrated, levels-of-processing effects are not dependent on processing time. For example, mean processing times for member responses to category and rhyme questions were virtually identical. However, all the preceding experiments have demonstrated that category member responses are recalled better than rhyme member responses.

GENERAL DISCUSSION

The major result in the present data is the repeated demonstration that, in free recall, subject-generated members of the set defined by the positive form of the question stem are not retrieved as well as nonmembers of that set. Further, the data clearly show that the nonmember items are not items that are better integrated with the question stem. If nonmember items were better integrated, cued recall would increase the superiority of those items. Instead, cued recall reverses the results, so that set members are recalled better than nonmembers. Taken together, these data provide strong evidence that the principle of congruity fails for subject-generated responses.

The data also point to one of the factors underlying the superior recall of nonmember items. Specifically, the reaction time data show that nonmember items take longer to generate, and thus presumably are more difficult to generate, than member items. Gardiner, Craik, and Bleasdale (1973) have shown that, with some qualifi-

Table 6
Mean Reaction Time in Seconds as a Function
of Conditions for Experiment 5

Question Type	Response Type		
	Member	Nonmember	Mean
Rhyme	2.00	2.53	2.27
Category	1.89	2.85	2.37
Mean	1.94	2.69	

cations, items more difficult to retrieve are better recalled on a subsequent trial. We suggest that the processes operating to produce the Gardiner et al. results may also have operated in the present experiments to produce, at least in part, the superiority of the nonmember items in recall.

A second outcome of some interest in these data is the demonstration that manipulation of the level of processing has the same effect on subject-generated responses as it has on experimenter-provided responses. That result is a nontrivial extension of previous research because the logic of the levels-of-processing hypothesis demands that the effect should hold whether the item characterized by the question stem is produced by the subject or by the experimenter. That is, processing at the level requested (e.g., semantic) must occur whether the subject is asked to generate an appropriate item or to determine whether a given item is appropriate. The current data, then, provide additional support for the levels-of-processing hypothesis.

In summary, level of processing and set membership (the principle of congruity) have both been shown to be strong determinants of recall for experimenter-provided items. The present experiments show that both of these variables also have a strong effect on recall of subject-generated items. However, though the effect of level of processing is the same for both types of items, the principle of congruity fails for subject-generated items.

REFERENCES

- BATTIG, W. R., & 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(3, Part 2).
- BUSCHKE, H. Two-dimensional recall: Immediate identification of clusters in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1977, 16, 201-215.
- CERMAK, L. S. *Human memory: Research and theory*. New York: Ronald, 1972.
- CRAIK, F. I. M., & LOCKHART, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 671-684.
- CRAIK, F. I. M., & TULVING, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, 104, 268-294.
- EPSTEIN, M. L., PHILLIPS, W. D., & JOHNSON, S. J. Recall of related and unrelated word pairs as a function of processing level. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 104, 149-152.

- GARDINER, J. M., CRAIK, F. I. M., & BLEASDALE, F. A. Retrieval difficulty and subsequent recall. *Memory & Cognition*, 1973, 1, 213-216.
- HYDE, T. S., & JENKINS, J. J. Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 471-480.
- SCHULMAN, A. I. Memory for words recently classified. *Memory & Cognition*, 1974, 2, 47-52.
- TILL, R. E., & JENKINS, J. J. The effects of cued orienting tasks on the free recall of words. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 489-498.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia University Press, 1944.

(Received for publication October 14, 1977;
revision accepted February 13, 1978.)