Journal of Experimental Psychology
1974, Vol. 102, No. 5, 778–787

RECALL AND RECOGNITION OF SEMANTICALLY ENCODED WORDS[1]

ENDEL TULVING[2]

Yale University
and
University of Toronto, Toronto, Ontario, Canada

The hypothesis was tested that the phenomenon of recognition failure of recallable words is attributable to the discrepancy between semantic properties of encoded target words and the semantic interpretation of corresponding recognition test words. Target words were presented for study and encoding in the presence of specific list cues, and subsequently tested with strong associates of target words serving as extralist cues. In one critical encoding condition, the list cues were semantically congruous with target words; in another, they were identical with target words. Results showed that congruous encoding rendered extralist cues only slightly less ineffective than incongruous encoding, and that under the conditions of "identical" encoding extralist cues were least effective. The results thus provided little support for the hypothesis, suggesting instead that specificity of encoding of word events in episodic memory transcends the semantic meaning of words.

This article describes two experiments designed to explore the phenomenon of recognition failure of recallable words recently reported by Tulving and Thomson (1973). In the Tulving and Thomson experiments, target words T (e.g., CHAIR) were presented at input in the company of cue words C (e.g., *glue*). Subjects expected to be tested for recall of T with C as a cue. When, instead of presenting C as a cue, the experimenter provided another word X (e.g., *table*)—which was a close semantic associate of the target word T but which had not appeared anywhere in the list—as an extralist retrieval cue, subjects could not readily use it as an aid in recall of T. Retrieval of the target word T was rather poor even in a subject-generated recognition test. In this test, subjects were instructed to write down free-association responses to the cue X, and then were asked to identify the generated responses as copies of target words from the studied list. Subjects had no difficulty generating copies of target words, but they did have difficulty recognizing them as "old" words from the studied list. In one experiment, for instance, only 24% of generated target words were correctly recognized in such a subject-generated recognition test, although 63% of all target words were subsequently recalled when the word C that had been paired with T at input was presented as cue.

The effectiveness of extralist retrieval cues consisting of strong associates of target words is a well-established phenomenon, under a variety of experimental conditions (e.g., Bahrick, 1969; Light, 1972; McLeod, Williams, & Broadbent, 1971; Thomson & Tulving, 1970), and hence the lack of their potency under the conditions of the Tulving and Thomson (1973) experiments is of some interest. What renders extralist cues ineffective? Why cannot certain recallable words be recognized?

A general interpretation of all acts of successful and unsuccessful retrieval of information from episodic memory is given by the encoding specificity principle (Tulving & Thomson, 1973). It states that the properties of the memory trace of a word event are determined by specific encoding

[1] This research was supported by National Science Foundation Grant 24171X. The article was written while the author was in residence as a Fellow at the Center for Advanced Study in the Behavioral Sciences in Stanford, California. Nora Keane helped with experimental work, Mary Pori with statistical analysis.

[2] Requests for reprints should be sent to E. Tulving, Department of Psychology, Yale University, New Haven, Connecticut 06510, or at the Department of Psychology, University of Toronto, Toronto M5S 1A1, Canada.

operations performed on the input stimuli, and that it is these properties, rather than the properties of the word in semantic memory, that determine the effectiveness of any given stimulus as a retrieval cue for the event. The principle suggests that if a stimulus in the retrieval environment renders possible or facilitates recall of the target word T, the retrieval information was appropriate to or compatible with the information contained in the episodic trace of T. Conversely, if a particular stimulus is ineffective in retrieving a particular trace, the conclusion follows that the appropriate relation was lacking. It is not yet known exactly what constitutes the "appropriate" relation between information contained in the trace and information present in the general retrieval environment or in a specific cue. Nor is the nature of the retrieval process as such clear. In absence of such knowledge, we assume that the unknown relation is one of "similarity" or "informational overlap," and that the retrieval process consists of some sort of combining of information from the two sources.

The failure of recognition of recallable target words in the Tulving and Thomson (1973) experiments thus can be interpreted as a special case of encoding specificity. It can be assumed that the trace of the target word T that had been encoded in relation to a particular list cue C did not possess sufficient overlap with the copy of the target word that was generated as a semantic associate of the extralist cue X. This is why X failed to bring about recovery of information in the trace of the target T. The encoding specificity principle, however, says nothing further about conditions or mechanisms responsible for inadequate overlap or complementarity between the trace of the C–T complex and the copy of T. But, if one accepts the reasoning so far, it is possible to ask the next question: Why do the encoding operations performed by the system on the target word T sometimes create a trace whose informational content cannot be matched by that of cue X, a strong semantic associate of T? Thus, for instance, why does the trace of the target word CHAIR, when presented as a part of the *glue*–CHAIR compound, differ from the trace of the target word CHAIR presented by itself, so that in the former case cues *table* and *chair* are relatively ineffective, although they both are quite effective under different experimental conditions (Bahrick, 1969; Light, 1972; McLeod et al., 1971; Thomson & Tulving, 1970; Tulving & Thomson, 1973)?

One hypothetical answer to this question was studied and evaluated in two experiments reported in this article. The answer consists of two parts: (*a*) what is stored, or at least what is retained after an interval, about a target word in a typical list-item memory experiment is the "meaning" of the word, and (*b*) pairing of the target word T with a certain list cue C at input produces a specific meaning of T that differs from the meaning of the same lexical unit perceived or produced in a different context, such as in a free-association test in which some other word X elicits T. Thus, CHAIR presented in a to-be-remembered list of words as a part of the *glue*–CHAIR compound is assumed to have a different semantic representation in memory than CHAIR generated as a semantic associate to the word *table*. The failure of recognition of recallable target words, the argument goes, then comes about for the same reason as the decrement in recognition of a homographic target word when the semantic interpretation of its copy at the time of the test is changed through changes in its verbal context (e.g., Light & Carter-Sobell, 1970; Winograd & Conn, 1971).

This semantic explanation of encoding specificity manifested in the phenomenon of recognition failure of recallable words was tested in the experiments described here. The experimental paradigm used was one in which the lexical identity of the target words was held constant and in which the effectiveness of various retrieval cues was examined as a function of semantic overlap between encoding context of the target word and the retrieval cue. Thus, retrieval of target word T was tested with extralist cue X following the presentation of the target word in one of three input contexts: (*a*) in the company of cue

words whose dominant semantic meaning was relatively incongruous with that of the target word T and the extralist cue X; (*b*) with cue words whose meaning was congruous with that of the semantic target word and the extralist cue; and (*c*) with cue words that were identical with target words. Thus, for instance, with the word CHAIR serving as target, the Incongruous encoding context or condition was represented by the pair *glue*–CHAIR, the Congruous condition by the pair *furnish*–CHAIR, and the Identical condition by *chair*–CHAIR. The extralist retrieval cue was *table*. It was assumed that the semantic information contained in the retrieval cue *table* would be most compatible with the stored information about the target word CHAIR presented in the *chair*–CHAIR compound (Identical encoding), somewhat less compatible with CHAIR presented in the *furnish*–CHAIR compound (Congruous encoding), and least compatible with information about CHAIR that had appeared as a part of the *glue*–CHAIR compound (Incongruous encoding). In keeping with the semantic encoding hypothesis it was expected that the effectiveness of the extralist cue would be directly related to the semantic similarity or compatibility between the cue and the target word.

EXPERIMENT I

Method

Materials and design. Target words presented in three different encoding contexts were recalled in three successive tests. The three encoding conditions were all represented within a single list that contained 24 target words. Each target word was accompanied at presentation by another word, a list, or input cue. Three types of list cues defined the three encoding conditions: (*a*) incongruous, (*b*) congruous, and (*c*) identical. The designations refer to the semantic relations between target words or subsequently presented extralist cues on the one hand and the list cues on the other. Each type of list cue appeared with 8 target words in the list.

Three successive retention tests were those used in previous experiments by Tulving and Thomson (1973): (*a*) a test in which strong extralist associates of target words served as retrieval cues, (*b*) a recognition test in which copies of target words were generated by *S*s as free associations to strong extralist associates of target words, and (*c*) a test in which list cues served as retrieval cues.

The basic set of materials used in the construction of experimental lists is presented in Table 1. Each of the 24 target words is shown together with its extralist associate which was used in the first cued recall test, and with both the Congruous and the Incongruous input cue. The 24 target words were specifically selected from the materials used in earlier experiments described by Tulving and Thomson (1973) to yield low recognition scores. Extralist cues and the incongruous input cues of these 24 target words were the same as those used in earlier experiments. The identical list cues are not shown in Table 1, for obvious reasons. Extralist and incongruous cues had been selected in the earlier experiments (Thomson & Tulving, 1970) from two sets of free-association norms. In these norms target words were primary associates of extralist cues; they were given as responses to incongruous cues 1% of the time, and extralist and incongruous cues were not related. The congruous input cues were generated by *E*, for the purpose of the present experiment, to represent words closely related semantically to both the target words and their extralist associates. By perusing the listing of the experimental materials displayed in Table 1, the reader can convince himself that the congruous list cues are more closely related to the target words and the corresponding extralist cue words than are the incongruous cues.

For the purpose of balancing of specific words and their encoding conditions, three different lists were constructed of the materials that appear in Table 1. The 24 target words were divided into three subsets of 8, each subset being used with a different set of input cues (incongruous, congruous, and identical) in one of the three lists. The lists were used with equal frequency in the experiment (specifically, since there were 50 *S*s, with 17, 17, and 16 *S*s respectively, in each group). Half of the target words in each subset of 8 were tested in the first retention test (with extralist associates of targets as cues), while the other half was tested in the second test (subject-generated recognition test), with the two halves counterbalanced across the two tests. Thus, each of the 24 target words was represented with equal frequency in all conditions of the experiment.

Prior to the presentation of the critical experimental list, *S*s were given a single practice and set-establishing list. It consisted of 24 target words, half of which were accompanied by input cues identical with target words and half accompanied by incongruous cue words comparable to those used in the critical experimental lists.

Subjects and procedure. Fifty *S*s, undergraduate students at Yale University and the Southern Connecticut State College, served in Experiment I, either in fulfillment of course requirements or for pay. They were tested in small groups of from 2 to 4 persons, in a session lasting approximately 40 min.

The initial instructions, given prior to the presentation of the set-establishing list, informed *S*s of the general nature of their task and the type of materials to be presented. They were told that they

would see pairs of words presented on the screen in front of the room, that their task was to remember the capitalized target words, but that they should also pay attention to the cue word accompanying each target word and notice the relation between them, since these cue words would help them to remember the target words. No special mention was made of the fact that some of the pairs would consist of identical words, nor were *S*s told anything particular about the retention tests.

The cue–target pairs from the set-establishing list were presented by means of a slide projector, at the rate of 3 sec/pair. At the end of the presentation *S*s were asked to open their recall booklets that had been handed out to them at the beginning of the session, turn to page 3 in the booklet, read the instructions, and proceed with the retention test. On this recall sheet the 24 input cues from the list were presented in an order unrelated to the input order. The instructions at the top of the recall sheet told *S*s to write down the capitalized words that had accompanied the given word cues in the list. Subjects were given 3 min. for the completion of this test.

Next, *S*s were told that they would be shown a new list of pairs of words, and that their task was the same as before—to remember each capitalized target word, noting the relation between it and its cue word. The nature of the relation between target and cue words was not specified or further commented upon.

The 24 pairs of the experimental list were then presented, again at the rate of 3 sec/pair. At the end of the presentation *S*s were asked to turn to page 6 of the recall booklet, read the instructions on the top of the page, and then proceed. (The pages for recall tests in the recall booklet were always separated by two numbered blank pages, hence this numbering system.) The recall instructions informed *S*s that their task was to try to recall the capitalized words they had seen in the list, but that this time the cues presented on the page were different from those they had seen in the input list. The *S*s were told to look at each cue word given on the recall sheet, see whether it reminded them of any target words they had just seen, and, if so, write it down beside the cue word. Twelve extralist cues, corresponding to four target words from each of the three encoding conditions, appeared on the recall sheet. Subjects were given 3 min. for this task.

Next, *S*s were asked to turn to page 9 in the recall booklet, which contained 12 extralist cues corresponding to the target words that had not yet been tested. Extralist cues, each followed by four spaces for free-association responses, were listed in a column on the left-hand side of the page. The instructions at the top of the page told *S*s that their task now was to produce free associations to the presented cue words. They were to write down four words in response to each of the cue words they saw on the page, words that the cue word "made them think of." This was an unpaced task, and *S*s were given as much time as they needed to complete it. When all *S*s in the group had finished writing, they were told to look at each word they had written down, decide whether or not it had occurred as one of the target (capitalized) words on the preceding study list, and, if they thought it had, draw a circle around the word. Four minutes were given for the completion of this task.

TABLE 1
TARGET WORDS AND CUES

Target words	Extralist cues	Congruous cues	Incongruous cues
BALL	tennis	player	whistle
BLUE	sky	heaven	pretty
CHAIR	table	furnish	glue
COAT	lining	cloth	covering
COLD	hot	fire	ground
DAY	night	evening	sun
DIRTY	clean	wash	barn
FLOWER	bloom	leaf	fruit
FOOD	eat	meal	moth
GREEN	grass	lawn	cheese
HAND	finger	arm	tool
HARD	soft	smooth	glass
HIGH	low	down	hope
LARGE	small	size	stomach
LIGHT	dark	black	head
MAN	woman	child	command
NEED	want	desire	bath
OPEN	closed	shut	country
ROUND	square	circle	cabbage
SHORT	long	thin	stem
SLOW	fast	speed	memory
WATER	lake	fish	whisky
WET	dry	moisture	cave
WIND	blow	move	noise

Finally, *S*s were asked to turn to page 12 of the recall booklet, read the instructions on the top of it, and proceed as instructed. This page contained all 24 input cues, in a scrambled order with respect to the order of their original appearance in the study list. The instructions informed *S*s that these were the input cues from the list, and that their task was to write down as many target (capitalized) words from the study list as they could, each one beside its corresponding cue word. Four minutes were allowed for this task.

The experimental session concluded with *E* explaining the purpose of the experiment to *S*s and answering any questions they had.

EXPERIMENT II

The *S*s in Experiment I seemed to do considerably better in the first retention test (cuing with extralist associates) than in the second (subject-generated recognition test). Since these data suggested that the order of the two tests may have been an important source of variance in the retention scores, Experiment II was con-

ducted to replicate Experiment I in every respect, with the sole exception of the order of the two tests. In Experiment II, the first of three tests was the subject-generated recognition test, while the second one was the cued recall test with extralist associates as cues. Otherwise the procedure was identical with that used in Experiment I in all details. The same materials were used both in the set-establishing list and the experimental list, the same counterbalancing procedure was followed, and the same final test with input cues was given. Thirty-six new *S*s from the same sources as in Experiment I served in this experiment.

Results

Set-establishing lists. The mean number of words recalled from the first set-establishing list was 16.7 in Experiment I, and 17.8 in Experiment II. In both experiments slightly more target words were recalled to identical than to incongruous cues (M = 8.8 and 9.5 in the two experiments, respectively). The mean number of intrusion errors was 1.34 and 1.83 per *S* in Experiments I and II, respectively. Of these intrusion errors in the two experiments, 44% and 58%, respectively, were "identical" intrusions: *S* gave a response identical with the cue although in the input list the cue had been paired with another word. It is possible that the higher recall of target words to identical cues was at least partly a consequence of inflation of recall scores by unidentifiable intrusions, but since it is not entirely clear what would constitute an appropriate correction, no attempt was made to assess the possible guessing bias.

Scoring and analysis. Analysis of the cued recall data, in all tests in which extralist or input cues were used, was straightforward. Each *S*'s recall score was the number of target words recalled, from a maximum of four or a maximum of eight per condition in the two tests, respectively. A stringent scoring criterion was used: *S* received credit for every target word only if it was correctly paired with its cue. (The inclusion in the data of target words recalled to incorrect cues, or to no particular cues, would not have changed the conclusions.)

Intrusion errors were ignored in all analyses. It is not immediately obvious how their inclusion in the results would change the overall conclusions drawn from the experiments.

In scoring the protocols from the subject-generated recognition tests, *S*s were given credit only for those copies of target words generated and recognized that belonged to the half of the input list being tested in the recognition test. In Experiment I, there was a total of 33 cases in the protocols of the 50 *S*s where a copy of the target word had been generated as a part of the free-association procedure and correctly recognized even though these words nominally belonged to the other half of the list that had been tested in the immediately preceding cued recall test using extralist retrieval cues. Since it was impossible to rule out the possibility that the words from the other half of the list constituted importations from the immediately preceding test rather than representing *S*s' memory for the material seen in the input list (particularly since most of these generated and recognized words had been in fact recalled by *S*s in the immediately preceding test), they were excluded from the analysis.

The data from the subject-generated recognition tests required a special treatment inasmuch as the maximum possible score of correct identification of target items depended upon the number of copies of target items any particular *S* had generated in the free-association test. There were three essential steps in the analysis of these data.

First, the number of copies of target items correctly generated in the free-association test were tabulated for each *S*, separately for target words from each of the three encoding contexts. The means of these scores were 2.84, 2.76, and 2.91 for the incongruous, congruous, and identical encoding conditions, respectively, in Experiment I, and 2.98, 3.11, and 2.44, for the same three encoding conditions in Experiment II. There is no obvious explana-

tion for the lower value of this statistic in the identical encoding condition in Experiment II, and since the same decrement was not observed in Experiment I, whatever implications this finding might have will be ignored in what follows.

The second step consisted of the calculation of proportions of copies of target words that were recognized by each *S* in each encoding context. Thus, for instance, if a given *S* generated three copies of the target item in a particular encoding condition, and recognized one of these, his recognition score (hit rate) was ⅓ or .33. The means of these proportions are provided as summary statistics for the subject-generated recognition test in Table 2. In a small number of cases, when *S* did not generate any copies of target words for a particular encoding condition in the free-association test, and hence could not possibly have recognized any, his score was considered indeterminate and was not entered into any analyses.

The third step in the analysis of the recognition scores was necessary for the purposes of the statistical evaluation of the data. Since the distributions of the recognition hit rates from individual *S*s in different encoding conditions in most cases were not distributed normally, statistical tests involving these measures were deemed inappropriate. Another measure of recognition performance was used instead. For each *S* in each encoding condition a right minus wrong (R − W) score was calculated.[3] Each target word whose copy had been generated and correctly recognized was considered R while each generated copy of a target word that was not recognized was considered W. For instance, an *S* who generated three copies of targets in a particular encoding condition, and recognized one of these, received an R − W score of 1 − 2 = − 1. The *differences* of these R − W scores for all three comparisons (incongruous vs. congruous, incongruous vs. identical, and congruous vs. identical) were distributed normally, thus permitting the use of *t* tests for related measures. Thus, while the data in Table 2 are summarized as mean hit rates, the corresponding analyses were conducted on the R − W scores as explained. There was, of course, a high although not a perfect correlation, within each of the encoding conditions, between individual *S*s' hit rates and R − W scores, the coefficient exceeding .90 in all cases.

Retrieval data. Table 2 presents a summary of the data on retrieval of target words from three different encoding conditions in three different tests.

Consider first the mean proportions of target words retrieved in various tests, with data pooled over the three encoding conditions in each test. Two observations are of interest. First, the order of the first two tests, both involving extralist cues, appears to have exerted considerable effect on the level of retrieval: *S*s' performance was higher in the first test (cued recall with extralist cues in Experiment I, and subject-generated recognition test in Experiment II) than in the second test. To evaluate test-order effects, a *t* test for unrelated measures was done comparing the differences between *S*s' scores on the cued recall and recognition tests in Experiment I with the same difference scores in Experiment II. The obtained value for *t* (84) was 4.57, $p < .01$. It appears, therefore, that some general test interference was produced in the course of the first test that suppressed the level of performance in the second. The proportions of copies of target items generated in the recognition test were identical at 71% in both experiments, suggesting that it is only the episodic and not the semantic (Tulving, 1972) component of the task that reflects interference, but given the design of the experiments, it is difficult to say anything more about the nature and sources of the test interference.

The second point with respect to retrieval scores pooled over different encoding conditions concerns the superiority of the performance in the final cued recall test over that in the other two tests, in both experiments. Statistical evaluation of the data was accomplished by means of *t* tests

[3] I am grateful to Perry Gluckman for suggesting this analysis.

TABLE 2

MEAN PERCENTAGES AND STANDARD DEVIATIONS OF TARGET WORDS RECALLED AND RECOGNIZED IN THREE TESTS AS A FUNCTION OF ENCODING CONTEXT

Experiment and encoding context	Test					
	Cued recall, extralist cues		Subject-generated recognition		Cued recall, list cues	
	M	*SD*	*M*	*SD*	*M*	*SD*
Experiment I ($N = 50$)						
Incongruous	51.5	34.0	26.3[a]	31.5	60.5	21.6
Congruous	52.0	27.1	36.5[a]	32.3	49.8	22.4
Identical	39.0	30.0	24.6[a]	31.1	46.5	26.1
M	47.5	23.4	28.6[a]	22.9	52.2	16.9
Experiment II ($N = 36$)						
Incongruous	28.5	29.4	42.4[b]	35.7	51.0	23.2
Congruous	36.1	29.6	46.1[b]	30.4	51.4	22.9
Identical	22.9	25.6	26.8[b]	27.2	36.8	30.3
M	29.2	20.1	39.9[b]	22.5	46.4	19.3

[a] $N = 45$.
[b] $N = 32$.

after the data in various tests were normalized in order to equate the three tests for maximum scores possible. The resultant t test was highly significant when the performance in the third test was compared with the performance in the second test in each of the two experiments, t (49) = 6.75 in Experiment I, and t (35) = 5.28 in Experiment II, both ps $< .01$. The t values were not significant when recall with input cues was compared with probability of retrieval in the first test in each of two experiments, t (49) = 1.74 in Experiment I, and t (35) = 1.88 in Experiment II, in both cases $.05 < p < .10$.

The data of primary interest concern differences in retrieval of target words that had been presented in different input contexts in the study list. The differences between encoding conditions were evaluated by means of t tests for related measures, separately for each of the three tests in both experiments. The results of these tests are tabulated in Table 3.

Although there was a slight numerical superiority in extralist cued recall and subject-generated recognition scores in the congruous input cue conditions over those in the incongruous input cue conditions, statistical analyses of the data, summarized in Table 3, showed that these differences did not reach acceptable degrees of reliability. The largest difference between the incongruous and congruous encoding conditions was observed in the subject-generated recognition test in Experiment I (26.3% vs. 36.5%), yielding a t (44) = 1.47, $.05 < p < .10$. In Experiment I, recall in response to list cues of target words that had appeared in the context of incongruous input cues was higher (60.5%) than recall of target words that had appeared in the context of congruous encoding cues (49.8%), t (49) = 2.80, $p < .01$, and this difference might be regarded as tempering the conclusion about lack of evidence for the effect of encoding conditions on retrievability of target words by extralist cues or their copies, but since Experiment II did not replicate this finding, its implications cannot be taken too seriously.

Since each to-be-remembered list word served as a target word in two successive tests, it was possible to examine recall or nonrecall of individual words in one test in relation to their recall or nonrecall in the other test. Indeed, the phenomenon of recognition failure of recallable words that provided the starting point for the present experiments can be most directly defined by demonstrating that subjects cannot

TABLE 3

SUMMARY OF t STATISTICS FROM TESTS COMPARING RECALL AND RECOGNITION SCORES BETWEEN PAIRS OF ENCODING CONDITIONS

Comparison	Cued recall, extralist cues		Subject-generated recognition		Cued recall, input cues	
	Experiment I	Experiment II	Experiment I	Experiment II	Experiment I	Experiment II
Congruous vs. Incongruous	.11	1.12	1.47	.78	−2.80**	.08
Incongruous vs. Identical	2.31*	1.10	.48	2.31*	3.10**	2.82**
Congruous vs. Identical	2.99**	2.45**	2.19*	2.68**	.88	2.60**

* $p < .05$.
** $p < .01$.

identify copies of target words as "old" but can produce these *same* words in presence of list cues. In the analysis of the "fate" of individual words that follows, retrieval of target words in extralist cue tests is compared with their retrieval in the final list-cue test.

Frequencies of words that are recalled in *both* the extralist cue test and the final list-cue test, as well as words that are recalled in *neither* of these two tests, are of little immediate interest. These data may reflect general "ease" or "difficulty" of words, breadth of encoding or failure of storage, effects of recoding or strengthening of traces in the first test, and other similar factors. Words that were retrieved only in the extralist cue test or only in the list-cue test provide more interesting data from the point of view of encoding specificity. These data are tabulated in Table 4.

Table 4 lists total frequencies of target words that were recalled or recognized in the extralist cue test but not recalled in the final list-cue test, and frequencies of target words not recalled or recognized in the extralist cue test but recalled in the final test. These two categories of words will be referred to as RN words and NR words, respectively, indicating the words' recall (R) or nonrecall (N) in the first (extralist cue) and the second (list cue) test. The left-hand panel of Table 4 presents data for target words initially tested in the extralist cued recall test, and the right-hand panel contains data for target words tested in the subject-generated recognition test.

Each entry in Table 4 consists of two figures. The first gives the total frequency (with the data pooled over all 50 *S*s in Experiment I and 36 *S*s in Experiment II) of RN words, while the second represents the total frequency of NR words. Each entry thus can be thought of as a ratio of two quantities, the number of RN words to the number of NR words, and the numerical value of this ratio, or some transformation of it, could be used as a rough index of the magnitude of the recognition failure of recallable words or the magnitude

TABLE 4

TOTAL FREQUENCIES OF TARGET WORDS RECALLED ONLY IN TEST WITH EXTRALIST CUES, IN RELATION TO FREQUENCIES OF WORDS RECALLED ONLY IN TEST WITH LIST CUES

Encoding condition	Recall with extralist cues			Subject-generated recognition test		
	Experiment I	Experiment II	Total	Experiment I	Experiment II	Total
Incongruous	17/45	15/54	32/99	10/47	17/26	27/73
Congruous	27/33	9/32	36/65	15/29	16/28	31/57
Identical	15/40	8/33	23/73	8/30	11/20	19/50
Total	59/118	32/119	91/237	33/106	44/74	77/180

of the failure of extralist cues in retrieval of words known to be represented in the episodic memory store.

Three features of the data in Table 4 should be mentioned. First, the data, pooled over both experiments, are rather similar in the two panels of Table 4, suggesting that underlying retrieval processes are probably not entirely dissimilar in extralist cued recall test and the subject-generated recognition test. Second, in all three encoding conditions, in both experiments and in both tests, the number of target words accessible only through list cues (NR words) was higher than that of words accessible only through extralist cues (RN words). Third, extralist cues were somewhat more effective for target words encoded in the congruous input condition than for those encoded under the other two conditions, as judged by the ratio of RN words to NR words, although the relatively small number of observations on which these data are based renders the differences statistically unreliable.

Discussion

The primary purpose of the experiments was to evaluate the hypothesis that specificity of encoding manifested in the phenomenon of recognition failure of recallable words (Tulving & Thomson, 1973) is a matter of specificity of semantic meaning of the to-be-remembered words and retrieval cues. Since in earlier experiments the target words were encoded in relation to list cues possessing little semantic similarity to target words and extralist cues, it seemed reasonable to argue that considerable semantic discrepancy existed between the target word as stored in that particular context and the literal copy of the target word presented as retrieval cue in the subject-generated recognition test. The graphemic units might have been the same in both cases, but they represented different bundles of semantic information.

The results of the present experiments do not provide much support for this semantic interpretation of encoding specificity. Some facilitation of retrieval of target words by extralist cues was observed under conditions where target words had been encoded in relation to cue words semantically much more compatible with both target words and extralist cues, in comparison with retrieval of target words encoded in relation to less congruous cues, but this facilitation was rather small and statistically not reliable. Furthermore, a considerable number of target words encoded in presence of congruous input cues were still not recognized in the subject-generated recognition test and could not be recalled to strongly associated extralist cues, even though they were recalled subsequently in presence of list cues, indicating that relevant information was available in the store. Finally, extralist cues were least effective, both in the recall and the subject-generated recognition tests, when target words were presented under the identical encoding condition where the input cue was nominally most compatible with the target word.

It is entirely possible that the hypothesis of semantic determination of encoding specificity fared so poorly in these experiments simply because the congruous encoding conditions were not congruous enough and the identical encoding conditions did not produce semantic encoding. If greater care were taken to ensure a high degree of semantic overlap between list cues and extralist cues the hypothesis might give a better account of itself. Nevertheless, the fact that it is not easy to eliminate the encoding specificity effect by manipulating the relation between input cues and extralist cues suggests that encoding of a familiar word in certain episodic memory tasks (Tulving, 1972) may entail degrees of specificity that cannot be achieved by changing semantic context of words. It may be more appropriate, therefore, to think about properties of memory traces of particular events and episodes as going beyond the properties of words as communicative units of language. It is not yet clear what these nonsemantic properties are.

The low level of recall and recognition of target words that had been presented in the input list under identical encoding conditions was quite unexpected. Several previous experiments, cited in the introduction, have shown that associated extralist retrieval cues are quite effective in facilitating recall of target words when these words occur in the study list without any manipulated context, and when the subjects are left free to encode them any way they want. There was no particular reason to expect, before doing the experiments described here, that the presentation of two identical copies of a word, one as "cue" and the other one as "target," would make any

substantial difference in the encoding of these list words. Yet the data were quite clear in showing that the Identical encoding condition was associated with lowest levels of recall and recognition, both in extralist cue tests and the final input cue test. It is this finding that appears to be most difficult to reconcile with the "semantic" interpretation of encoding specificity, suggesting that other or at least additional factors are involved.

The data from the practice list, in which immediate recall of target words was just about the same under identical encoding and incongruous encoding conditions, suggest that it was not the initial level of storage of "identically" encoded words that was impaired but rather that the low level of subsequent recall was some consequence of events that took place after the study of the list. It is difficult to say anything definite about this impairment on the basis of existing data. It may be that subjects encoded target words under identical conditions primarily in terms of their phonetic properties, in terms of a clang association between the cue and the target word, and that this information was lost more rapidly than the specific "semantic" information, or that the subsequently presented retrieval cues were somewhat less appropriate for providing access to this kind of information. It is also conceivable that the low level of retrieval of target words under identical conditions was a consequence of a high degree of similarity of encoding operations performed on these target words during the presentation of the list. If part of the information stored in the memory trace of an event represents the encoding operation performed on the input (Tulving & Thomson, 1973), then a high degree of similarity of encoding operations performed on a number of list items could render the resulting traces less unique and hence somehow less readily retrievable (Lesgold & Goldman, 1973; von Restorff, 1933). Be it as it may, the present data rather clearly suggest that the identity relation between two items, one designated as cue and the other as target, does not represent the extreme position on the dimension of associative or semantic relatedness.

Finally, in relation to earlier experiments (Tulving & Thomson, 1973) the present data showed rather a high degree of effectiveness of extralist retrieval cues, despite the fact that the materials for the present experiment were specially selected from among those used in earlier studies to yield a minimal level of retrieval in presence of extralist cues. It was hoped that such minimal levels of retrieval would have made it easier to demonstrate the effect of a congruous semantic input context. It may be that the use of a study list containing different sorts of cue–target relations was responsible for the discrepancy between the present and earlier experiments with respect to effectiveness of extralist cues, although for the time being this must remain pure speculation.

REFERENCES

BAHRICK, H. P. Measurement of memory by prompted recall. *Journal of Experimental Psychology*, 1969, 79, 213–219.

LESGOLD, A. M., & GOLDMAN, S. R. Encoding uniqueness and the imagery mnemonic in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 193–202.

LIGHT, L. L. Homonyms and synonyms as retrieval cues. *Journal of Experimental Psychology*, 1972, 96, 255–262.

LIGHT, L. L., & CARTER-SOBELL, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 1–11.

MCLEOD, P. D., WILLIAMS, C. E., & BROADBENT, D. E. Free recall with assistance from one and from two retrieval cues. *British Journal of Psychology*, 1971, 62, 59–65.

THOMSON, D. M., & TULVING, E. Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, 1970, 86, 255–262.

TULVING, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

TULVING, E., & THOMSON, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, 80, 352–373.

VON RESTORFF, H. Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung*, 1933, 18, 299–342.

WINOGRAD, E., & CONN, C. P. Evidence from recognition memory for specific encoding of unmodified homographs. *Journal of Verbal Learning and Verbal Behavior*, 1971, 10, 702–706.

(Received July 9, 1973)