

Recognition failure of recallable famous names in a hybrid semantic-episodic memory task

DAVID G. PAYNE and JAMES H. NEELY
Purdue University, West Lafayette, Indiana 47907

Subjects studied a list of lowercase cue and uppercase target pairs containing famous names (e.g., *betsy ROSS*) and nonfamous names (e.g., *edwin CONWAY*). Following recognition tests in which surnames were tested in the absence of their cues, cued recall tests for the surnames were given. In semantic recognition and recall tests, the response to a surname was to be made solely on the basis of its fame, whether or not it had appeared in the study list. In episodic memory tests, the response to a surname was to be made solely on the basis of whether or not it had appeared in the study list, regardless of its fame. For individual subjects, the type (i.e., episodic vs. semantic) of recognition test and cued recall test was always different. The observed recognition failure rate for the studied famous names tested in the semantic recognition and episodic cued recall condition was significantly lower than that predicted by the Tulving-Wiseman (1975) function, whereas the recognition failure rate for the episodic recognition and semantic cued recall condition was significantly higher than predicted. The data are interpreted as providing some support for the distinction between episodic and semantic memory.

Under certain conditions, subjects are unable to recognize a word as having been previously studied even though they are later able to recall it as being a previously studied word. This phenomenon has been termed "recognition failure of recallable words" (Tulving & Thomson, 1973). Tulving and Wiseman (1975) have shown that this recognition failure (RF) can, under a wide range of study and testing conditions, be accurately predicted by the following quadratic equation:

$$PRF = 1 - [RN - .5 (RN - RN^2)], \quad (1)$$

where PRF is the predicted probability of RF and RN is the overall probability of recognition. (This quadratic equation will hereafter be referred to as the TWF, for Tulving-Wiseman function.)

RF has been interpreted by some theorists as supporting Tulving's (1972) episodic-semantic memory distinction. According to these theorists, RF depends upon the formation of a "unique" episodic memory trace (Kintsch, 1974, pp. 77-82; Watkins & Tulving, 1975, p. 6). However, this interpretation has been challenged by studies demonstrating that the TWF accurately predicts RF in a semantic memory task in which the test items have not been previously studied in the experimental context (e.g., Muter, 1978).

The research reported here was supported by Grant R01 HD15054 awarded to the second author by the National Institute of Child Health and Human Development. Requests for reprints should be addressed to James H. Neely, Department of Psychological Sciences, Purdue University, West Lafayette Indiana 47907.

In a replication and extension of Muter's (1978) study, Neely and Payne (1983) directly compared episodic and semantic RF rates for famous names in a single experiment. In this experiment, following presentation of a study list containing both famous and nonfamous names, subjects received either episodic or semantic recognition and cued recall tests for the studied and nonstudied surnames. In the episodic memory tests, subjects were to base their responses on whether or not the test name had appeared in the study list, regardless of its fame. In the semantic memory tests, subjects were to respond on the basis of whether or not the test name was the name of a famous person, regardless of whether or not it had appeared in the study list. Results showed that the TWF accurately predicted RF (1) for the famous surnames in the semantic memory test, regardless of whether or not they had been previously studied, and (2) for the studied famous surnames, regardless of whether the memory test was an episodic or a semantic memory test.

The procedure used in the present experiment was identical to that of Neely and Payne (1983), with the exception that the type of test (episodic or semantic) was varied, rather than held constant, across the recognition and cued recall tests. That is, subjects who received an episodic surname recognition test received a semantic surname recall test, and vice versa. The issue of primary interest is whether the TWF can also accurately predict RF in this hybrid task. If it can, the data would call into the question the idea that episodic and semantic memory tests necessarily tap different kinds of information. If the TWF cannot accurately predict RF in this hybrid task, the results would indicate that, at

least in the RF paradigm, episodic and semantic memory tasks tap different information, through differences in either storage and/or retrieval processes.

METHOD

Subjects

Eighty-two undergraduate introductory psychology students were tested in groups of 16-23 each.

Materials

The practice list consisted of a random ordering of 12 weakly related and 12 unrelated cue-target word pairs. The two critical lists (List A and List B) consisted of a random ordering of 24 name items (12 famous and 12 nonfamous) and 24 nonname cue-target word pairs (12 weakly related and 12 unrelated). All nonname items were taken from Tulving and Thomson (1973), and all famous names were the names of people who had achieved fame prior to 1950. The famous and nonfamous surnames were common names, as measured by frequency of occurrence in telephone directories. An arbitrary first, or first plus middle, name(s) (hereafter called a first name) and a fictitious character description were generated for each of the nonfamous names (see Neely & Payne, 1983, for more details and a complete listing of all name items). The critical lists were constructed by randomly assigning half of the famous and nonfamous names to serve as study items in List A and as lures in List B, and vice versa. The same nonname cue and target items appeared in both critical lists, with assignment of targets to related and unrelated cue-target pairs counterbalanced across lists.

All study items were typed and prepared as slides. The practice list test and the four critical list tests were prepared in booklet form. The practice list cued recall test consisted of the 24 cue words presented in the practice list. The critical list tests were presented in the following order: nonname recognition, name recognition, nonname cued recall, and, finally, name cued recall. (Since the first and third critical list tests involved the nonname items, these tests will not be described here. The data from these nonname tests are reported in Neely and Payne, 1983.) The name recognition test contained 48 names (24 famous and 24 nonfamous). The name cued recall test contained 48 (24 famous and 24 nonfamous) character-description-plus-first-name cues. On both the recognition and cued recall tests, for individual subjects, half of the famous and nonfamous items were from the study list and half were not.

Procedure

The practice and critical list study and testing procedures were identical to those used by Neely and Payne (1983), with the exception that in the present experiment, for individual subjects, the type of task (i.e., episodic vs. semantic) was varied across the name recognition and name cued recall tests (see Neely & Payne, 1983, for further details concerning all study and testing procedures).

All items were presented at a 5-sec/item rate. Before the practice list, subjects were informed that they would be shown slides containing two words, one in capital letters and one in lowercase letters, and that their task was to try to remember the words in capital letters (i.e., target words). They were also told to pay attention to the words in lowercase letters (i.e., cue words), since these words might help them remember the capitalized words. After the list items were presented, subjects were given 3 min to complete the practice list cued recall test.

After the practice list test, subjects were told that half of the slides (nonname items) in the second (critical) list would be similar to the slides presented in the first list and that the remaining slides contained the names of real or fictitious people. Subjects were told that for these name items, their memory for the surnames would be tested and that the first name might help them to remember the surnames.

Following critical list presentation, subjects were given four tests for the critical list items. Subjects were given 5 min to complete each test. For the episodic name recognition test, subjects were instructed to circle a Y if they believed the surname on the test sheet had appeared in the study list and to circle an N if they felt it had not. Subjects given the semantic name recognition test were to base their answers on whether or not they recognized the surnames as being the names of people who achieved fame prior to 1950. It was emphasized that if the subjects recognized a surname as being one that appeared in the study list but not as being famous, then N was to be circled.

For the episodic name cued recall test, subjects were told to write down beside each of the first names that had appeared in the list the surname that had appeared with it in the list. They were also told that if they were not able to recall the surname when they read the first name, they could read the description presented along with that first name as a further retrieval cue. It was emphasized that if the description made the subject think of a surname that was not presented in the slides, then that surname was not to be written down. Subjects receiving the semantic cued recall test were told to write down beside a first-name-plus-description cue the surname of the famous person who corresponded to that cue. Once again, it was emphasized that recall was not to be limited to only those famous surnames presented in the study list and that if a surname from the study list was not famous it should not be written down.

In summary, all subjects took episodic recognition and cued recall tests for the nonname items. For the name items, subjects in the episodic-semantic condition took an episodic recognition test followed by a semantic cued recall test, whereas subjects in the semantic-episodic condition received a semantic recognition test followed by an episodic cued recall test. Subjects were tested in four large groups formed by factorially crossing the two testing conditions (episodic-semantic and semantic-episodic) with the two critical lists (Lists A and B).

RESULTS

For the semantic recognition test, false alarms for the famous names from the study list were false-positive recognitions of nonfamous names from the study list. For the episodic recognition test, false alarms for the studied famous names were false-positive recognitions of famous names not from the study list (see Neely & Payne, 1983, for more details concerning the scoring of the name recognition and cued recall tests).

Overall Cued Recall and Recognition Performance

The mean proportion of recognition hits, false alarms, correct cued recall, and cued recall intrusions for the semantic-episodic condition were .71, .33, .58, and .07, respectively. The corresponding values for the episodic-semantic condition were .78, .16, .66, and .15, respectively. A 2 (condition: episodic-semantic vs. semantic-episodic) by 2 (list: A vs. B) by 2 (retention test: correct cued recall vs. recognition hits) mixed-factor analysis of variance was performed on the data from the famous names from the study list. All effects called statistically significant had p values of $<.05$.

As expected, subjects' recognition hit rates exceeded their correct cued recall rates [$F(1,78) = 35.6$, $MSe = .018$], although the magnitude of this effect was different for the two study lists [$F(1,78) = 15.4$, $MSe = .018$, for the List by Retention Test interaction]. Also, subjects in the episodic-semantic condition had both higher

recognition hit rates and higher correct cued recall rates than did subjects in the semantic-episodic condition [$F(1,78) = 4.4$, $MSe = .042$]. The only other significant effect was the effect of list [$F(1,78) = 4.4$, $MSe = .042$].

Recognition Failure

The data from the famous names in the study list were used to construct a 2 by 2 (recalled/not recalled by recognized/not recognized) contingency table for each subject. The first four columns of Table 1 present these contingency tables aggregated across subjects. The fifth and sixth columns of Table 1 give the observed and predicted RF rates for the aggregate data, along with an indication of statistical significance. (The binomial distribution was used to evaluate statistical significance; see Neely & Payne, 1983, for details.)

Since RF rates for different items may not be independent for a particular subject, the independence assumption underlying the binomial distribution analysis of the aggregate data may have been violated. Therefore, in a second analysis, we computed observed and predicted RF rates for each subject and used the individual subjects, rather than the individual items, as our unit of analysis. (Because RF is based on recalled items, the data from those few subjects who recalled no famous names from the study list were ignored in the individual subject analysis.) A sign test was used to determine whether the number of subjects whose observed RF fell above or below the predicted RF was statistically greater than would be expected by chance. Columns 7 and 8 of Table 1 give the observed and predicted RF rates averaged across those subjects who provided data for that condition. Column 9 in Table 1 presents the number of subjects, of the total number providing data for that condition, whose RF rates fell above the predicted RF rates.

As can be seen in Table 1, the observed RF rate was significantly lower than predicted in the semantic-episodic condition and significantly higher than predicted in the episodic-semantic condition. This pattern of results was obtained for both the aggregate data and the individual subject data. The lower than predicted RF in the semantic-episodic condition and the higher than predicted RF in the episodic-semantic condition cannot be explained simply by the fact that the overall recognition rates, used to predict RF, were different in these

two conditions. This conclusion is supported by two different pieces of evidence. First, in a post hoc analysis, we equated the two conditions on overall recognition hit rates by discarding the data from the subjects with the highest and lowest recognition hit rates in the episodic-semantic and semantic-episodic conditions, respectively. Even then, a sign test using individual subjects as the unit of analysis indicated that the observed RF rates in these two conditions still deviated significantly, and in opposite directions, from those predicted by the TWF. Second, within the episodic RF paradigm, Flexser and Tulving (1978) have shown that the TWF accurately predicts RF over a much larger range of recognition hit rates than that observed in the present experiment. Hence, if in the present experiment the only difference between the two conditions was the .07 difference in overall recognition hit rate, then one would expect, based on Flexser and Tulving's data, that the observed RF should still be accurately predicted by the TWF.

DISCUSSION

The fact that the observed RF rates for the episodic-semantic and semantic-episodic conditions deviated significantly, and in opposite directions, from those predicted by the TWF indicates quite clearly that RF is not independent of the type (i.e., episodic or semantic) of memory queries used in the RF paradigm. As long as the type of memory task is held constant across the recognition and cued recall tests, the TWF accurately predicts both episodic RF and semantic RF (Neely & Payne, 1983). The present study demonstrates, however, that when the type of task is varied across the recognition and cued recall tests, the observed RF rates deviate significantly from those predicted by the TWF. The present data thus indicate that, at least for the recognition and cued recall tests used in the RF paradigm, episodic and semantic memory tasks tap different underlying information, through differences in either storage and/or retrieval processes.

It is also interesting that the observed RF rates for the two conditions employed in the present experiment not only deviated from those predicted by the TWF but did so in opposite directions. That is, in the semantic-episodic condition the observed RF rate was lower than predicted, whereas in the episodic-semantic condition it was higher than predicted. At present, we have no firm explanation to offer as to why the RF rates from these two conditions deviated in opposite directions from the predicted rates. Rather, we simply point to this result as a further bit of evidence that, in the RF paradigm at least, episodic and semantic memory tasks do not tap identical information.

In summary, the present data make it clear that while semantic

Table 1
A 2 by 2 Classification of the Recall and Recognition of the Famous Names in the Study List Along With the Aggregate and Individual-Subject Observed and Predicted Recognition Failure Rates

Condition	Recognized		Not Recognized		Aggregate Data		Individual-Subject Data		#O<P
	R	NR	R	NR	O	P	O	P	
Semantic-episodic (N = 38)	228	96	33	98	.13	.19*	.15	.20	28/37*
Episodic-Semantic (N = 44)	282	128	65	52	.19	.14*	.20	.15	12/39*

Note—R = recalled, NR = not recalled, O = observed recognition failure (RF), P = predicted RF; #O<P = number of subjects with observed RF lower than predicted RF out of the total number of subjects with nonzero cued recall scores. A very few subjects failed to respond to all items on the recognition tests. As a result, the total number of observations does not equal a multiple of 12.
 * $p < .05$.

RF rates are analogous to those obtained in the standard episodic RF paradigm (Muter, 1978; Neely & Payne, 1983), the storage and/or retrieval processes underlying performance in the episodic memory and semantic memory conditions are not identical. If these processes were indeed identical, then the RF rates for the episodic-semantic and semantic-episodic conditions employed in the present experiment would have been accurately predicted by the TWF, as were the RF rates for the episodic-episodic and semantic-semantic conditions in Neely and Payne's (1983) experiment. Since the observed RF rates in the present experiment were not well predicted by the TWF, these data provide some support for the episodic-semantic memory distinction.

REFERENCES

- FLEXSER, A. J., & TULVING, E. Retrieval independence and recall. *Psychological Review*, 1978, **85**, 153-171.
- KINTSCH, W. *The representation of meaning in memory*. Hillsdale, N.J: Erlbaum, 1974.
- MUTER, P. Recognition failure of recallable words in semantic memory. *Memory & Cognition*, 1978, **6**, 9-12.
- NEELY, J. H., & PAYNE, D. G. A direct comparison of recognition failure rates for recallable names in episodic and semantic memory tests. *Memory & Cognition*, 1983, **11**, 161-171.
- 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.
- TULVING, E., & WISEMAN, S. Relation between recognition and recognition failure of recallable words. *Bulletin of the Psychonomic Society*, 1975, **6**, 79-82.
- WATKINS, M. J., & TULVING, E. Episodic memory: When recognition fails. *Journal of Experimental Psychology: General*, 1975, **104**, 5-29.

(Received for publication January 18, 1983.)