Recognition Failure of Recallable Unique Names: Evidence For an Empirical Law of Memory and Learning

Lars-Göran Nilsson
University of Umeå, Umeå, Sweden

Janine Law and Endel Tulving
University of Toronto, Toronto, Canada

An experiment is described in which subjects ($N = 106$) studied unique names of famous individuals, such as ISAAC NEWTON and GEORGE WASHINGTON, and unique geographic names, such as TORONTO and STOCKHOLM, in the context of descriptive phrases. In a subsequent recognition test of the names in the absence of their study context, subjects failed to recognize many names that they could recall in the presence of the study context. These results (a) demonstrate the generality of the phenomenon of recognition failure of recallable words, (b) limit Muter's (1984) claim that unique names of famous people constitute an exception to the Tulving-Wiseman (1975) function, and (c) provide further support for an empirical law concerning the relation between recognition and recall.

Empirical laws in the science of memory and learning are notable mainly by their absence (Cohen, 1985). Probably because of the scarcity of relations between variables that hold independently of particular conditions of observations, the dominant zeitgeist is one of skepticism regarding general "truths" regarding phenomena of learning and memory. Thus, many students of memory and learning seem to have accepted as inevitable the ubiquity of interactions in the outcomes of their experiments (e.g., Jenkins, 1979) and have adopted a "sour grapes" attitude toward general principles (e.g., Baddeley, 1979, p. 359; Mandler, 1979, p. 305). Given such a milieu, the appearance of a persistent regularity in experimental findings and the possibility that such a regularity represents a "law" that holds under a wide variety of conditions should be a happening of interest.

In this article we discuss the possibility that a regularity of law-like proportions is found in the outcomes of experiments conducted under the rubric of "recognition failure of recallable words." We also present the results of an experiment demonstrating that the regularity holds for a particular type of to-be-remembered items, namely, unique names of famous people and well-known places, that have previously been declared to lie outside the boundaries of the regularity (Muter, 1984). By empirically removing an apparently conflicting fact, we strengthen the case for the putative "law."

In a "recognition-failure" experiment, subjects study to-be-remembered words, each of which is presented in a particular verbal context, and then attempt the retrieval of the to-be-remembered words in two successive tests. The first test is one of recognition in which copies of the to-be-remembered words appear without their study context; the second is a cued-recall test in which the intralist contexts of the to-be-remembered (or target) words are presented as retrieval cues. In this situation, subjects frequently recall study-list words that they failed to identify as "old" on the preceding recognition test, thus demonstrating the phenomenon of recognition failure of recallable words, or, simply, recognition failure.

Recognition failure is a robust phenomenon. It has been obtained with a variety of to-be-remembered items (e.g., Neely & Payne, 1983; Nilsson, Dinniwell, & Tulving, 1987; Tulving & O. C. Watkins, 1977), with different kinds of contextual materials (e.g., Bartling & Thompson, 1977; Park & Tulving, reported in Tulving, 1983), in experimentally naive subjects (e.g., Tulving & Thomson, 1973) as well as those thoroughly familiar with the phenomenon and the paradigm (Rabinowitz, Mandler, & Barsalou, 1977; Wiseman & Tulving, 1975), with elderly people as well as young ones (Rabinowitz, 1984), with both yes–no and forced-choice recognition tests (M. J. Watkins & Tulving, 1975), with different kinds of distractor items in the recognition test (e.g., M. J. Watkins & Tulving, 1975) or even no distractors at all (Nilsson et al., 1987; Wallace, 1978), with and without lure cues in the cued-recall test (e.g., Tulving & O. C. Watkins, 1977; M. J. Watkins & Tulving, 1975), over retention intervals measured in minutes (e.g., Begg, 1979; Rabinowitz, 1984) as well as weeks (Donnelly, 1988), and in situations in which recall is higher than recognition (Wiseman & Tulving, 1976) as well as those in which recognition is higher than recall (e.g., Postman, 1975).

The Tulving-Wiseman Function

All of the foregoing variables affect recall or recognition or both. But they leave largely invariant the relation between recognition conditionalized on recall and overall recognition

hit rate. This invariance is reflected in the close adherence of data from a large number of recognition-failure experiments to a simple quadratic function described by Tulving and Wiseman (1975), as follows:

$$P(Rn|Rc) = P(Rn) + .5[P(Rn) - P(Rn)^2], \qquad (1)$$

where Rn refers to recognition hits, and Rc refers to cued recall. The probabilities of recognition failure, $P(\overline{Rn}|Rc)$—the complement of recognition conditionalized on recall in Equation 1—observed in experiments show relatively little deviation from the probabilities "predicted" by the function.

Note that the particular expression of the Tulving-Wiseman function—in terms of the relation between overall recognition and recognition conditionalized on recall—is simply one of several possible variants of the underlying relation between *recognition* and *recall*. The value of the single constant $c = .5$ in the function is an index of the dependency between recognition and (cued) recall in a 2×2 contingency table (Tulving & Wiseman, 1975; Tulving, 1983, chap. 13). Complete independence between recognition and recall would be represented by $c = 0$ in the function, and complete dependency by larger values of c, whose upper bound varies inversely with the recognition hit rate. The degree of dependency between recognition and recall represented by $c = .5$ could be described as "moderate."

The fact that the data from many experiments closely adhere to the Tulving-Wiseman function has prompted Jones (1983, 1984) to label the function a "law." And Cohen (1985) has pointed to the theoretical linkage between his own third law of memory and the Tulving-Wiseman law. A tentative statement of the "law" could be put as follows: If a to-be-remembered item of information T is encoded with respect to some other "contextual" item of information C, then recognition of T in the absence of C and the cued recall of T in the presence of C show moderate dependency that remains largely invariant with the variables that otherwise affect recognition and recall.

What makes this "law" interesting is not just the systematic relation between overall recognition and recognition failure: Countless experiments in psychology have yielded data showing a variable Y to be related to another variable X, yet there has been no felt need to refer to these relations as "laws." Rather it is the *invariance* of the relation between *recognition* and cued *recall* that the function describes that distinguishes the recognition-failure findings from many other sets of data in research on memory. A discussion of the generality of the phenomenon of recognition failure of recallable words and a summary of a large number of experiments whose results conform to the Tulving and Wiseman (1975) function can be found in Tulving (1983). Many variables have been identified that greatly affect either recall or recognition, or both (e.g., Anderson & Bower, 1972; Brown, 1976; McCormack, 1972; Gillund & Shiffrin, 1984), yet they have little or no effect on the *relation* between the two. Given that even in mature physical sciences the type of law that is "most eagerly sought" is one "which says in effect 'this function of these variables under given conditions is always *constant*'" (Holton & Brush, 1973, p. 182), the invariance captured by the Tulving-Wise-

man function may be regarded as a modest success story in our developing field of research. The function asserts that the relation between recognition and cued recall, under given conditions, is always largely the same regardless of the values of these measures of retention or their determinants.

The Problem

No regularity of nature identified by the methods of science holds absolutely; there are always exceptions to any law (Holton & Brush, 1973). An important research problem is to study such exceptions and their relations to the general law. Certain apparent exceptions to the Tulving-Wiseman function have also been reported (Tulving, 1983, p. 289–290). A few experiments whose designs have conformed to the recognition-failure paradigm have found either no recognition failure at all or they have found considerably less recognition failure than expected by the Tulving-Wiseman function. Because the existence of these exceptions has a clear bearing on the status of the regularity of the recognition-failure findings as an empirical "law," it is of some importance to examine the exceptions and to seek their experimental and theoretical clarification. We are engaged in a research program aimed at such clarification (Nilsson & Tulving, 1986; Nilsson et al., 1987). The present article describes a part of it.

The exceptions to the Tulving-Wiseman function known at this stage of research seem to fall into two broad categories. Exceptions in one category come about because they violate the encoding conditions specified in the "law"; exceptions in the second category occur because the retrieval conditions are not met. Consider the two categories in turn. Experiments that form the first category are represented by those reported by Bartling and Thompson (1977), Begg (1979), Gardiner and Tulving (1980), and Neely and Payne (1983). Exceptions in these cases have been interpreted by Gardiner and Tulving (1980) in terms of insufficient association of the to-be-remembered item with its list context and by Neely and Payne (1983) as a poor integration of the two items of each pair. According to the statement of the putative "law," as given in the second paragraph of this article, the conditions under which it holds include the requirement that the to-be-remembered information T be "encoded with respect to" the context information C. Gardiner and Tulving's interpretation thus reduces to the hypothesis that exceptions to the Tulving-Wiseman function occur if one of the critical conditions of the "law" is not fulfilled.

A second category of exceptions is represented by experiments reported by Nilsson and Shaps (1980, 1981), in which categorized word pairs were used as study materials. This exception has been explained by Nilsson et al. (1987) as a consequence of a situation in which the other critical condition—that the retrieval information in the recognition and cued recall be unrelated—is not fulfilled: The *functional* retrieval information in the recognition test includes the information provided in the cued-recall test, and thus recognition of T is not tested in the absence of C.

Another exception to the function is represented by the outcome of an experiment by Muter (1984). It has not yet been critically evaluated, and it is not known, therefore,

whether it fits into one of the two tentatively established categories. In Muter's (1984) experiment, subjects were tested as to how well they could recognize and subsequently recall unique names of well-known people (such as Kierkegaard and Ataturk) in a *semantic-memory analogue* of the typical episodic recognition-failure experiment. In this analogue, there is no study list. Subjects are simply given two successive tests, a "recognition-of-fame" test and a "produce-the-famous-name" test. Other experiments have shown that the data from such semantic-memory analogues of recognition-failure do conform to the Tulving-Wiseman function when commonly occurring names of famous people (e.g., Cooper, Davis, Ross) are used as target items (Muter, 1978, 1984; Neely & Payne, 1983). Muter (1984), however, also used "unique" names (e.g., Ataturk, Kierkegaard, Rachmaninoff) as target items and found no recognition failure: Subjects could always recognize the names that they subsequently recalled to cues such as *First president of the republic of Turkey: Kemal* ____ and *Danish philosopher, founder of existentialism: Soren* ____.

Thus, Muter's results constitute a glaring exception to the Tulving-Wiseman function. Muter (1984) also claimed that his findings provided support to a version of the venerable generation/recognition theory that had been proposed during the early days of research on recognition failure as an explanation for the phenomenon (Martin, 1975; Reder, Anderson, & Bjork, 1974). According to this version of the theory, recognition failure can occur only for words that have multiple meanings, represented by multiple nodes in the network of (permanent) memory, and cannot occur if the to-be-remembered item is coded by a single node. Muter (1984) assumed, with Anderson (1977), that each unique name is represented by a single node in semantic memory. Because, according to theory, this node must be activated in both the recognition and the recall tests, each name that can be recalled must also be recognized.

Purpose of the Experiment

Muter's (1984) findings diminish the generality of the data that define the putative "law." His theoretical interpretation, too, is in conflict with the "law" as tentatively formulated: The formulation specifies the conditions under which the "law" holds in terms of encoding processes and the relation of retrieval cues in recognition and recall, but says nothing about the importance of type of target items or the number of nodes coding them. Under these conditions, it seemed worthwhile to check on the generality of Muter's (1984) findings, in a conventional episodic recognition-failure experiment. In the study phase of the experiment, subjects saw the full names of famous people, such as GEORGE WASHINGTON, CHARLES DARWIN, and ALFRED NOBEL in the context of descriptive phrases characterizing the individuals concerned. In the recognition test, they had to decide whether each of the test names, presented in the absence of the study context, had appeared in the study list. In the subsequent cued-recall test, they had to produce each name in the presence of the previously seen descriptive phrase. In addition to the names of famous people, in an attempt to extend the generality of our

findings, we also investigated recognition failure for well-known unique geographical names (e.g., NEW YORK, TORONTO and STOCKHOLM).

If Muter's (1984) findings can be generalized to episodic memory, and if his arguments concerning the number of hypothetical nodes coding an item and the relevance of this factor to recognition failure are valid, then our experiment should yield an outcome comparable to his. If, on the other hand, Muter (1984) obtained his results for reasons other than the type of the target items used, then recognition failure of recallable unique famous names may occur and its magnitude may conform to the Tulving-Wiseman (1975) function.

Method

Plan of the Experiment

All subjects were treated identically, although different subgroups were tested with different sets of counterbalanced materials.

At study subjects saw a series of phrases and unique person names (e.g., *He was the first of a long line but the only one on horseback—* GEORGE WASHINGTON). In the recognition test, subjects saw target names (e.g., GEORGE WASHINGTON) together with distractors, items consisting of names of famous people not seen in the study list (e.g., CHARLES DARWIN). In the recall test, subjects were shown cues consisting of the phrases from the study list (e.g., *He was the first of a long line, but the only one on horseback:* ____), together with comparable phrases fitting the names presented as distractors in the recognition test (e.g., *A famous scientist of uncertain ancestry:* ____).

A set of geographical names was also presented to each subject for study, with subsequent recognition and recall tests, under the same conditions. At study the subjects saw a set of phrase–name pairs (e.g., *A well-known building for music in* VIENNA), at recognition they saw the names of both study-list and distractor items (e.g. VIENNA, BORDEAUX) without phrases, and at recall they responded to the phrases that had accompanied study-list items (e.g., *A well-known building for music in* ____) as well as phrases corresponding to the geographical names that had served as distractors in the preceding recognition test (e.g., *Made of grapes in* ____).

The basic question of interest had to do with the extent to which, if any, recognition failure of recallable names would occur with these unique names of famous people and unique geographical names.

Materials and Design

The pool of materials consisted of 64 pairs of phrases and corresponding names of famous persons (e.g., *He made scientists work harder:* ALFRED NOBEL; *The world's most famous piano player and mathematics teacher:* ALBERT EINSTEIN), and 64 pairs of phrases and corresponding geographic names (e.g., *Building for clergyman in* CANTERBURY; *Hat first made in* MOROCCO). Names selected as target items had to be unique and well known. The phrases to accompany the target names were constructed to satisfy two requirements: (a) They had to make some sense in terms of what the subjects were expected to know about the famous individuals or geographic places named, and (b) they had to preclude ready generation of the target name in the absence of specific study.

Two parallel study lists were compiled from this pool. Each list contained two blocks of study pairs: 32 pairs of phrases and person names, and 32 pairs of phrases and geographic names. The 64 pairs of each study list appeared in a booklet consisting of three pages. The

first 16 pairs and the last 16 in the booklet were filler items. They were studied by the subjects and included in the recognition test only, but the data were neither tabulated nor analyzed. The 32 pairs in the middle of the study booklet (16 pairs with a person name and 16 pairs with a geographic name) constituted the critical pairs that were tested in both recognition and recall. They provided the data of the experiment.

The recognition-test booklet contained all 128 names from the original pool. Thus, for each subject there were 64 names from the study list (32 person names and 32 geographic names) and 64 comparable distractor names not previously presented to that subject. The 128 names in the recognition test appeared in a booklet of two pages with separate blocks of 32 person names and 32 geographic names on each page. Targets and distractors appeared in a random order in each of the two groups of names on each page. A 6-point scale (−3 to +3) appeared next to each name for the subjects' confidence ratings of their recognition responses.

The recall-test booklet also consisted of two pages, with 32 phrases for person names appearing on one page and 32 phrases for geographic names on the other. Sixteen phrases on each page had appeared as critical study-list pairs, whereas 16 phrases, fitting the distractor names from the recognition test, had not been previously seen by the subject in the experiment.

Subjects

A total of 106 students at the University of Toronto participated in the experiment, which was conducted as part of an introductory cognitive psychology course in two successive class sessions 1 week apart. Students in the class were given a study trial at the beginning of the session on the first occasion, and the recognition and cued-recall tests at the beginning of the session 1 week later. The 106 students who were present on both occasions provided the data for the experiment.

Procedure

Subjects studied the 64 study-list pairs by reading each phrase–name pair and rating the meaningfulness of each pair. Before commencing, they were told that they were participating in a "memory experiment." They were instructed to study the pairs in the booklet in preparation for a test in which the phrase of each pair would be presented as a retrieval cue for the recall of the name. As a part of the study procedure they were to rate each pair with respect to the "sense" that it made, by using a 3-point scale on which a rating of 2 indicated that the phrase made perfect sense in relation to the name, 1 indicated that the phrase made some sense, and 0 indicated that the phrase did not make any sense in relation to the name. Subjects were allowed to study and rate the pairs in the study booklet at their own pace. The experimenter collected the study booklets when all students had finished the study/rating task, 10 min after the beginning of the task.

Seven days after the study trial a recognition test was given, followed by a cued-recall test. The subjects recorded their responses in the recognition-test booklet, which listed 64 names of famous people and 64 geographic names. They were to indicate for each name whether they remembered having seen it in the study list a week earlier; they were also asked to rate their confidence in each decision made, on a 3-point scale. If they were "absolutely certain" that they had seen the name during study they were to circle +3 on the 6-point scale appearing next to each name, if they were "reasonably certain" they were to circle +2, and if they were guessing that they had seen the name at study they were to circle +1. Comparable negative ratings were to be given to test names the subjects judged to be new: If they were "absolutely certain," "reasonably certain," or "guessing" that they had not seen a given name they were to indicate

this by circling −3, −2, and −1 respectively. The recognition test was unpaced; the slowest subject took 11 min to complete it

In the cued-recall test, which immediately succeeded the recognition test, subjects were given a booklet with 64 phrases, 32 for people's names and 32 for geographical names. They were told that their task was to write down the names fitting the descriptive phrases that they remembered from the study list, or failing that, to write down names seen in the recognition test, or any other names, that seemed to belong with the fragments. Subjects were also asked to indicate the "source" of each response they made. If they thought the response corresponded to a name seen in the study list, they were simply to list the name. If they thought the response corresponded to a name seen only in the recognition test, they were to mark their response with an R (for recognition test). And if they did not remember the response from either the study list or the recognition test, but could produce a name that seemed to "go with" the phrase nevertheless, they were to mark it with a G (for general knowledge, or guessing). The recall test was also unpaced; the slowest subject needed 15 min for its completion.

Results

Before we consider the main findings of interest, some preliminary observations need to be made. One such observation concerns the ability of subjects to respond correctly to the cue phrases in the recall test in the absence of the study trial. Given the plan of our experiment, such responding would be possible not only because of occasionally correct guessing on the basis of general (semantic) knowledge, but also because of recognition-test priming, that is, on the basis of the subject's encounter with the target names (only) in the recognition test. The data from the main part of the experiment are interesting to the extent that recognition and recall are determined by the study episode, rather than by guessing and test priming. It is necessary, therefore, to ascertain that recognition and recall were affected by the study list, and to estimate the extent to which our subjects could have responded correctly in the cued-recall test on the basis of either their general knowledge or priming in the recognition test.

The issue of recognition-test (priming) effects on cued recall in the recognition-failure paradigm has been explicitly dealt with in previous publications (e.g., Bowyer & Humphreys, 1979; Donnelly, 1988; Flexser & Tulving, 1982; Humphreys & Bowyer, 1980; Postman, 1975; Tulving & O. C. Watkins, 1977). The overall conclusion that has emerged from this research can be summarized in the following three statements. (a) Priming effects do occur: testing target items for recognition does enhance their subsequent cued recall (Humphreys & Bowyer, 1980); (b) these effects occur for both recognized and unrecognized target items (Donnelly, 1988); and (c) the overall contribution of such priming effects to the phenomenon of recognition failure is relatively small (Flexser & Tulving, 1982). The comparison of cued recall for target items and recognition-test distractors in our experiment provides some relevant information concerning the potential influence of semantic guessing and recognition-test priming on cued recall of both recognized and unrecognized items. We summarize these data next.

Our 106 subjects, in the cued-recall test for the names of famous persons, produced a total of 159 correct responses to the phrases that they had *not* seen in the study list although they had seen these names as distractors in the immediately

preceding recognition test. (In the cued-recall tests, we gave subjects credit for recall of the name of a famous person even if they put down only the surname of the person, and we ignored spelling errors.) The subjects thought that they had remembered 63 of these 159 names from the recognition test, that they had produced 63 responses from their general knowledge, and had seen 33 of the names in the study list. The corresponding figures for geography names were smaller: There were 35 responses, consisting of 13 identified as "recognition-test" items, 16 as "general-knowledge" items, and 6 (mistakenly) identified as study-list items. Because there were 1,696 opportunities (106 subjects × 16 responses) for making correct responses on the basis of person names encountered in the recognition tests, and the same number of opportunities for geography names, the figures just presented represent proportions of .09 for person names and .02 for the geography names. As these proportions include both names guessed on the basis of semantic knowledge and names that were primed in the recognition test, they represent the upper-limit estimate of the effect of recognition-test priming.

Cued recall of study-list names was considerably higher. Our subjects produced at total of 664 person names and 193 geography names, or 39% and 11%, respectively. Of these responses, 85% and 71% for the two categories of names, respectively, were identified by the subjects as having been seen in the study list. Of the person name responses correctly recalled, 12% were identified as having been seen in the recognition test and 3% of the correct responses were generated from general knowledge. For geography names these values were 12% and 17%, respectively.

Thus, although the overall levels of cued recall were rather low, especially for geography names, and although subjects did assign rather large proportions of correct responses to sources other than the study episode, the bulk of the data we will report does represent episodic retention, that is, the consequences of the subjects' learning of the study list. In the main analysis of the data, we ignored subjects' statements as to the "source" of their knowledge about each name they produced on the cued-recall test. We did so for two reasons. First, one of our major objectives is to compare our data with those reported in all other recognition-failure experiments; in these other experiments subjective "source" judgments were not collected. Second, the data indicate that subjects' statements regarding the "source" of their knowledge are not too reliable to begin with.

To sum up these preliminaries: Semantic guessing and recognition-test priming do inflate the probabilities of cued recall, but they account for only a relatively small proportion of cued recall of recognized and unrecognized target names. The bulk of any recognition failure observed in the experiment can therefore be assumed to occur independently of the recognition-test priming.

Analysis of the Data

The basic unit of analysis in this experiment is a single subject item. Each subject item was assigned to one of four possible outcome categories, depending upon whether it was judged to be "old" or "new" in the recognition test, and whether it was produced or not produced to the cue phrase

in the subsequent recall test. The frequencies, or proportions, of subject items in the cells and margins of the resulting 2 × 2 contingency table, and the measures derived from them, such as the conditional probability of nonrecognition given recall (the "recognition-failure" measure), constitute the data of interest.

We report the results of overall analyses of this kind. We also report the results of subsidiary analyses, based on various subject-determined subcategories of items and responses. Recall that (a) at study, subjects rated each cue–target pair for its *meaningfulness*, and (b) at recognition, they expressed their *confidence* about each recognition judgment. This latter measure can be regarded as a criterion of recognition decision (Murdock, 1974, p. 27). With three subjective categories of meaningfulness of study pairs and three criteria of recognition decision, it is possible to examine the 2 × 2 contingency data for 9 (3 × 3) subsets of study-list items. We report the main data in this form. It should be noted that subsets of items classified by study-list meaningfulness are mutually exclusive, whereas those classified by the recognition criterion are not.

All summary data appear in the form of *weighted* proportions. That is, the frequencies in the cells of a 2 × 2 contingency table are summed over all subjects, and probabilities of recognition and recall, as well as recognition failure measures, are calculated on the basis of these sums.

Person Names

Table 1 presents a summary of the person-name data, classified by study-list meaningfulness ratings and, within each rating category, by three recognition criteria. As there were three missing observations, the frequencies of observations in the three meaningfulness categories add up to a total of 1693 rather than 1696. The data in the top row in each meaningfulness category represent the outcome of the analysis for the "lenient" recognition criterion. In this analysis, an item is considered as having been correctly recognized if the subject called it "old" with a confidence rating of +1, +2, or +3. The data in the second row represent the outcome with an "intermediate" criterion. In this case, "recognized" items are those assigned confidence ratings of +2 or +3, and "nonrecognized" items are those assigned confidence ratings of +1, -1, -2, or -3. The data in the third row in each meaningfulness category are based on a "strict" criterion: Only those items with confidence judgments of +3 are regarded as "recognized," and items with all other confidence ratings are classified as "not recognized."

For each subset of observations, Table 1 shows the recognition hit rate, probability of recall, nonrecognition conditionalized on recall ("recognition failure"), the level of recognition failure expected on the basis of the Tulving-Wiseman function, and the deviation of the observed recognition failure from the expected one.[1] This latter measure, the *deviation*, is

---

[1] The cell entries in the 2 × 2 contingency table corresponding to each set of data can be calculated from the information given in Table 1. For instance, the proportion of items not recognized *and* recalled is given by the product of recall and recognition failure. Other entries can then be calculated by subtraction.

Table 1

Recognition Hit Rate (Rn), Probability of Recall (Rc), and Probability of Recognition Failure for Unique Person Names

Decision criterion	Rn	Rc	Recognition failure		
			Observed	Expected[a]	Deviation
Low meaningfulness (412 observations)					
Lenient	.612	.155	.234	.269	.035
Intermediate	.422	.155	.422	.456	.034
Strict	.245	.155	.609	.663	.054
Medium meaningfulness (543 observations)					
Lenient	.753	.357	.180	.154	−.026
Intermediate	.573	.357	.304	.305	.001
Strict	.359	.357	.485	.526	.041
High meaningfulness (738 observations)					
Lenient	.770	.553	.164	.141	−.023
Intermediate	.638	.553	.289	.247	−.042
Strict	.450	.553	.468	.426	−.042

[a] On the basis of the Tulving-Wiseman function.

of primary interest in this experiment: It describes the extent to which the observed data conform to the Tulving-Wiseman function and thus the extent to which they are indistinguishable from recognition-failure data obtained with target words with many meanings that in Muter's (1984) theory are represented by multiple nodes.

Table 1 shows, first, that both recall and recognition increased with rated meaningfulness of study pairs. It also shows that the recognition hit rate decreased with the increasingly stringent criteria, a necessary consequence of the analytical method used. The same was also true for false alarms, which were .35, .22, and .10 for lenient, intermediate, and strict criteria, respectively.

The important data shown in Table 1 concern the probabilities of nonrecognition conditionalized on recall, and the deviations of these probabilities from the Tulving-Wiseman function. These data, shown by filled circles, are graphically depicted in Figure 1, which makes it possible to appreciate the overall picture at a glance. In keeping with tradition, the ordinate of the graph in Figure 1 represents recognition conditionalized on recall, that is, the complement of the recognition failure measure. The amount of recognition failure as shown in Table 1 is indicated by the distance of the data points from the top of the graph.

Figure 1 shows that for all nine subsets of observations, the deviations from the function were small, well within the range of variability found in typical recognition-failure experiments. Although systematic relations between deviations and meaningfulness ratings, and deviations and recognition criteria, may be discerned in Table 1, the effects are small and of little consequence in the broader picture with which we are concerned. The important finding is that recognition failure of recallable unique names of famous people does occur, and that the amount of this recognition failure is comparable to that observed in many previous experiments in which other kinds of to-be-remembered words have been used (e.g., Tulving, 1983, Figure 13.1). Even more compelling is the finding

that rated meaningfulness of the to-be-remembered materials at study, which does have a strong effect on recall and a noticeable effect on recognition, leaves intact the largely invariant *relation* between recall and recognition described by the Tulving-Wiseman function (see Equation 1). Similarly interesting is the finding that the relation between recall and recognition of unique names remains largely invariant with drastic changes in the decision criterion in recognition, changes that resulted in large variations in the recognition hit rate: There was close agreement between observed recognition failure and that expected on the basis of the Tulving-Wiseman function for all three levels of criterion, in each of the three meaningfulness categories.

Geography Names

Table 2 presents a summary of the data for geography names in the same format used in Table 1 for person names. The total number of observations represented in Table 2 is 1,691, because there were five missing observations.

The pattern of simple recall and recognition probabilites of geography names shown in Table 2 is very similar to that of person names, although the absolute levels of both recall and recognition are lower. Both recall and recognition increased with rated meaningfulness of study-list materials, and both recognition hit rates and false alarm rates decreased with increasing stringency of the recognition criterion. The false alarm rates for lenient, intermediate and strict criteria were .38, .18, and .05, respectively.

The important data are again the observed probabilities of recognition failure, and the extent to which these probabilities conform to the Tulving-Wiseman function. In order to appreciate these data fully in comparison with those obtained for person names, we have depicted them in Figure 1 as open circles. Six of the data points for geography names conform
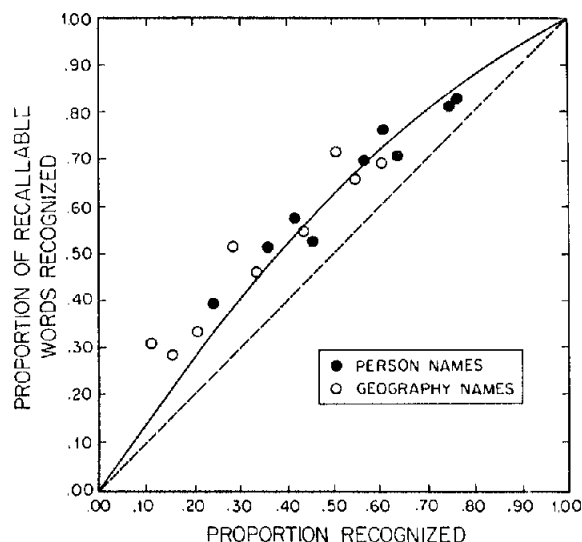

Figure 1. Proportion of recallable words recognized as a function of the proportion of recognition of all target words (original data).

Table 2

Recognition Hit Rate (Rn), Probability of Recall (Rc), and Probability of Recognition Failure for Unique Geography Names

Decision criterion	Rn	Rc	Recognition failure		
			Observed	Expected[a]	Deviation
Low meaningfulness (832 observations)					
Lenient	.501	.047	.282	.374	.092
Intermediate	.281	.047	.487	.618	.131
Strict	.111	.047	.692	.840	.148
Medium meaningfulness (504 observations)					
Lenient	.546	.117	.339	.330	−.009
Intermediate	.339	.117	.542	.549	.007
Strict	.155	.117	.712	.780	.068
High meaningfulness (355 observations)					
Lenient	.611	.259	.304	.270	−.034
Intermediate	.437	.259	.446	.440	−.006
Strict	.206	.259	.663	.712	.049

[a] On the basis of the Tulving-Wiseman function.

to the function in much the same way as all nine data points did for person names. These six points represent the data for all three decision criteria in recognition of study pairs that were rated as medium or high in meaningfulness. Thus, recognition failure of recallable unique geography names occurs to the same extent as that for unique names of famous people provided that the phrase–name pairs are reasonably meaningful for the subjects at study. When, on the other hand, rated meaningfulness is low, the observed amount of recognition failure adheres less closely to the Tulving-Wiseman function. For all three recognition criteria in the low meaningfulness category the deviation from the function is noticeable; as seen in Figure 1 these data points fall above the function, indicating less recognition failure than expected.

Homogenization of Data

The qualitative and quantitative characteristics of the data from this experiment agree well with the expectations based on the results of many other recognition-failure experiments summarized by the Tulving-Wiseman function: Recognition and (cued) recall exhibit a moderate degree of positive dependency in the contingency analysis. Any dependency between measures in a 2 × 2 contingency table is composed of three components: subject correlations and item correlations, and between the two tests, subject-item interactions (Flexser, 1981). The contributions of each of the three components to the dependency can be estimated by a procedure that Flexser (1981) has described. The procedure, which Flexser has dubbed "homogenization", results in a 2 × 2 contingency table in which any dependency reflects only the effect of subject-item interactions. In the homogenized data, intersubject and interitem variances are zero because all subjects' scores correspond to the mean subject, and all item scores to the mean item. We homogenized the data from our experiment in order to find out to what extent the observed de-

pendency between recognition and recall reflected subject and item correlations.

We report the results of the homogenization procedure for 12 subsets of data, corresponding to combinations of two separate lists of materials with three levels of recognition criterion, for each of two types of target names, persons and geography. Because the standard homogenization procedure requires the raw data provided by n subjects tested on two successive tests with the same set of m items, each test outcome being recorded as 1 or 0 for success or failure, it was not possible to homogenize the data for subsets defined by study-list meaningfulness ratings.

The results of the homogenization procedure applied to the 12 subsets of data are shown in Table 3. Table 3 gives simple probabilities of recognition and recall for each of the 12 subsets, together with both the "raw" and "adjusted" (homogenized) recognition-failure probabilities, $P(\overline{Rn}|Rc)$. Because the homogenization procedure leaves the marginals of the 2 × 2 tables unchanged, the probabilities of overall recognition and recall in Table 3 correspond to those shown in Tables 1 and 2, as do the "raw" recognition-failure probabilities. As before, the cell entries can be calculated from the data provided. Because the number of subjects tested with the two lists was slightly different, List 1 proportions are based on 880 observations each (55 subjects × 16 items), and List 2 proportions on 816 observations each (51 subjects × 16 items).

The important data in Table 3 are the adjusted recognition-failure measures, that is, conditionalized probabilities of recognition miss given recall based on homogenized 2 × 2 tables. Not surprisingly, all 12 adjusted probabilities are larger than the corresponding raw probabilities, indicating decreased dependency between recognition and recall. The whole picture can be seen at a glance in Figure 2, which shows homogenized recognition conditionalized on recall as a function of overall recognition.

Figure 2 shows that, with a single conspicuous exception, for which we have no better explanation than measurement error, the homogenized recognition-failure scores vary highly systematically with the overall recognition hit rate. The data points adhere closely to the Tulving-Wiseman function although they tend to lie below it, showing somewhat greater amounts of recognition failure than expected according to the function.[2] A comparison of Figure 2 with Figure 1 also shows that the variability (scatter) of data points around the Tulving-Wiseman function is smaller for the homogenized data.

A specific finding concerns the fact that the smaller-than-normal recognition failure with low-meaningful geography names (the large deviation scores in Table 1) seems to have been attributable to large subject and especially large item correlations between the two tests for those subsets of data. For List 2, shown in Table 3, for instance, item correlations

---

[2] As a rough index of the effect of the homogenization on the dependency between recognition and recall, we can calculate the c constants of the Tulving-Wiseman function for both the raw and adjusted recognition-failure scores in Table 3. The mean c of the 12 raw scores was .83, the mean c of the 12 homogenized scores was .37. Flexser (1981, Table 2) has reported the results of a homogenization analysis for the data from an experiment reported by Wiseman and

Table 3
Results of Flexser's (1981) Homogenization Procedure Applied to 12 Subsets of Data

| | | Person names | | | | Geography names | | |
| | | | $P(\overline{Rn}|Rc)$ | | | | | $P(\overline{Rn}|Rc)$ | |
Subset	$P(Rn)$	$P(Rc)$	Raw	Adjusted	$P(Rn)$	$P(Rc)$	Raw	Adjusted
				List 1				
Lenient	.706	.368	.179	.232	.555	.144	.315	.336
Intermediate	.548	.368	.296	.355	.355	.144	.551	.583
Strict	.390	.368	.454	.514	.152	.144	.748	.780
				List 2				
Lenient	.745	.420	.172	.199	.518	.077	.302	.501
Intermediate	.583	.420	.318	.350	.320	.077	.349	.590
Strict	.348	.420	.516	.584	.126	.077	.571	.787

were .59, .80, and .82, for the three levels of criterion, respectively, whereas subject correlations were smaller, ranging from .28 to .54. Homogenization treatment, which removes the effects of these correlations, brought the data for all subsets of geography names into good agreement with the data for person names.

Discussion

The results of this experiment have demonstrated that recognition failure of recallable unique names of famous people, as well as unique names of well-known geographical locations, occurs to an extent indistinguishable from that observed in many other experiments using different kinds of target items for recognition and cued recall. The data for both kinds of names, and for subsets of items defined in terms of rated meaningfulness of study materials and in terms of recognition criteria, closely adhered to the Tulving-Wiseman (1975) function, thereby demonstrating an invariant relation between recognition and cued recall. The results of the homogenization analysis confirmed previous findings (Flexser, 1981, Table 2) in showing that the dependency between recognition and recall is considerably reduced and that the probability of recognition failure is increased when the effects of subject and item differences are statistically eliminated, but they did not alter the substance of the overall findings of such invariance.

Thus, the major empirical contribution made by the research described here concerns the generality of the recognition-failure phenomenon. The conclusion drawn earlier by Tulving and O. C. Watkins (1977) that words with single meanings do not constitute an exception to the Tulving-Wiseman function has been corroborated by our results. It

Tulving (1975). According to our calculations, the mean c of the raw recognition-failure scores from eight separate conditions in that analysis was .58, and the mean c of the homogenized scores was .38. Thus, roughly speaking, it looks as if homogenization reduces the dependency between recognition and recall, indexed by c, by something like a factor of two.

does not seem to matter whether these single meaning words are of the type used by Tulving and O. C. Watkins (1977), or whether they are unique famous names of particular people or specific geographic locations. The phenomenon of recognition failure, its conformity to the Tulving-Wiseman function, and the invariance of the relation between recall and recognition reflected in the function seem to hold generally for common and unique names, as well as for many other kinds of target items.

This fact is worth knowing and emphasizing, because many writers seem to think differently. Independently of Muter's (1984) findings and conclusion, Horton and Mills (1984), in a recent critical survey of human learning and memory in the *Annual Review of Psychology*, referred to the recognition failure phenomenon as an important contribution with valuable implications for the understanding of memory, but they also expressed concern about the generality of the phenomenon across materials. Recent textbook writers (e.g., Best, 1986;
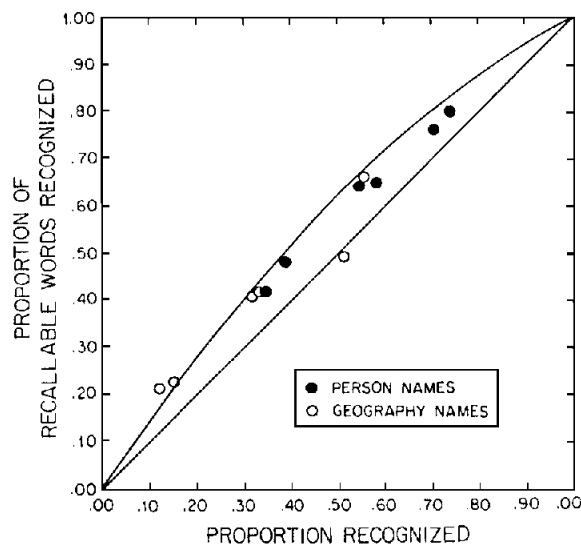
Figure 2. Proportion of recallable words recognized as a function of the proportion of recognition of all target words (homogenized data).

Stern, 1985) make the same point: The recognition failure phenomenon lacks generality across different types of materials. Others have noted that the effects of semantic context in episodic recall and recognition are reduced for words with a single meaning (e.g. Gillund & Shiffrin, 1984; Reder et al., 1974). Our findings serve as yet another reminder that these concerns are without foundation in fact.

In discussing these results, we raise four issues: (a) the theoretical interpretation of our findings, (b) the interpretation of Muter's (1984) results, (c) the implications of our findings for the generation/recognition theory of recall, and (d) the status of the Tulving-Wiseman function as an empirical "law."

The Flexser-Tulving Model

Our results fit well into the Flexser and Tulving (1978, but see also Flexser & Tulving, 1982) theory of recognition failure. The theory, based on the encoding specificity principle, has been presented in the form of a mathematical model that entails no fixed parameters. It accounts not only for the occurrence of recognition failure but also for its magnitude, that is, the Tulving-Wiseman function and its single empirical constant $c = .5$.

According to the theory, a unique episodic trace is formed at the time of study, composed of features of both the cue and target items. The semantic-memory status of these items is important only to the extent that the appropriate semantic information is used by the system in the construction of the trace. Otherwise this status—whether specified in terms of number of nodes in a network or some other way—is irrelevant. Retrieval is successful to the extent that the retrieval information matches the specifically encoded information. Recognition failure of recallable words occurs, to the extent specified by the Tulving-Wiseman function, whenever the retrieval information provided at the recognition test is independent, or nearly so, of that available to the subject at the recall test. If, on the other hand, the retrieval information provided by the recall cue constitutes a subset of the information in the recognition cue, or if the two sets of retrieval information completely overlap, then all recalled words are necessarily recognized and there is no recognition failure.

The condition of largely independent retrieval information was satisfied in our experiment in the sense that, in the absence of relevant episodic information, the target could not have been readily predicted from the cue, nor the cue from the target. The descriptive phrases that we used as cues were constructed with this purpose in mind. We also assume that the condition of independent retrieval information has been fulfilled in all the other recognition-failure experiments that have yielded data conforming to the Tulving-Wiseman function. As we mentioned in the introduction, however, the condition was not satisfied in the series of experiments by Nilsson and Shaps (1980, 1981) that showed no recognition failure with study lists consisting of pairs of items in which category names were cues and same-category instances were targets. In this situation, subjects can readily predict or generate the recall cue (category name) when given the recognition cue (an instance of the category). When the same

materials were used under mixed-list conditions where the predictability of the cue from the target was greatly reduced, recognition failure again occurred (Nilsson et al., 1987).

The Flexser-Tulving (1978, 1982) model says nothing about the relevance of *type* of to-be-remembered information for recognition failure. The fact that in our experiment recognition failure was observed with unique person and geography names is thus unsurprising from the point of view of the model. But what about Muter's (1984) results? Why did he obtain normal amounts of recognition failure with common names of famous people and fail to observe any recognition failure with his set of unique names?

Muter's Findings

In the absence of facts from appropriate direct comparisons we can only speculate on the answer to this question, leaving a more decisive answer to be provided by future experiments. We think that a subtle confounding occurred in Muter's (1984) experiment between the type of name (unique versus common) and the amount of overlap of retrieval information available to the subjects in the two tests. Specifically, we assume that when Muter used unique names, the recall cue provided little or no information in addition to that already presented to the subject in the previous recognition test, whereas with common names, the retrieval information in the recall test exceeded that in the previous recognition test. If these assumptions are true, then, according to the Flexser-Tulving model (1978, 1982), recognition failure should occur with common names but not with unique names. This difference in the outcome does not result from differences in the type of to-be-remembered names but from the differences in the relation between the retrieval information in recognition and recall for the two types of names.

Let us illustrate this speculative explanation with a specific concrete example.

A University of Toronto student who knows little about Turkey's history is likely to respond negatively to both the name Ataturk and the cue, *first president of republic of Turkey: Kemal _____*. On the other hand, a student who does know that Ataturk is a famous name and makes the positive response in the "recognition" test probably also knows that Ataturk was a Turkish statesman. That information, even if not a part of the nominal cue provided to the subject in the "recognition" test, is thus implicitly or "functionally" present (cf. Nilsson et al., 1987) when the subject recognizes Ataturk's name as a famous one. When that information is explicitly or nominally present in the subsequent cued-recall test, it represents only a subset of the total functional retrieval information already used by the subject in the preceding "recognition" test. As a consequence, it cannot effect retrieval of information that was not retrieved in the recognition test: There is no recognition failure.

The situation is different with respect to common names as studied not only by Muter (1978, 1984) but also by Neely and Payne (1983). With these names, the retrieval information in the cued-recall test exceeded that available to the subjects in the recognition test. The subject may not recognize Cooper or Ross as famous names, not because he or she does not

know the famous people with these names, but because the information provided for the successful solution of the task is insufficient. When more effective retrieval information is provided, e.g., *American author of "The Last of the Mohicans: James Fenimore* _____, or, *The Maker of the American Flag: Betsy* _____, the subject can make more effective use of his or her stored knowledge about literature and history. In other words, Cooper and Ross are not famous names, whereas James Fenimore Cooper and Betsy Ross are. Had Muter used these full names, he probably would have found as little fame-recognition failure for them as he did for names such as Ataturk, Kierkegaard, and Rachmaninoff.

Our interpretation of Muter's (1984) results is admittedly post hoc, but it is a relatively simple matter to test it experimentally. Pending such a test we simply note that if our basic assumptions regarding Muter's (1984) experiment are true, then the general model of Flexser and Tulving (1978) satisfactorily accounts for his findings. And if so, Muter's (1984) results represent an instance of our second category of exceptions to the function that we mentioned earlier—exceptions attributable to the testing of the recognition of T in the presence of C.

Quite independently of our speculative interpretation of Muter's (1984) findings, our results make clear that Muter's (1984) findings must be attributable to reasons other than the use of unique names. Our results also have implications for the version of the generation/recognition theory that Muter used to interpret his results. We consider this next.

Generation/Recognition Theory

Muter (1984) interpreted his findings in terms of a revised version of the theory (Martin, 1975; Reder et al., 1974). According to this theory, in the recognition test "a node in memory is automatically accessed, and if it bears the relevant information a positive response is emitted" (Muter, 1984, p. 201). Successful recall depends on (fallible) access to the node and on the presence of requisite information in it. Muter assumed (a) that recognition failure occurred for common names, because common names are represented in memory by multiple nodes, and (b) that it did not occur for unique names, because these are represented by a single node. When a common name is examined by the subjects in the "recognition" test, only some of its multiple nodes are activated. The associated information concerning the fame of the named person need not be available in the activated nodes and hence the subject makes a negative fame decision. In the subsequent recall test, a different subset of nodes may become activated. If one of these nodes bears the relevant information, the subject may now produce the name of the famous person, and "recognition failure" has occurred. The situation is different with unique names. A person with a unique name is represented in semantic memory by a single node which must be activated for successful retrieval in both the recognition and the recall tests. Consequently, recognition failure of the recallable unique name cannot occur.

The multiple-sense and multiple-node generation/recognition theory can account for Muter's (1984) particular findings, but it is at variance with ours. Because our results leave little

doubt that recognition failure occurs with names such as GEORGE WASHINGTON, ISAAC NEWTON, TORONTO, and STOCKHOLM, the conclusion must be drawn *either* that the number of meanings or semantic senses of a word or a name is irrelevant for the phenomenon of recognition failure *or* that even unique names have multiple senses. In the former case, the multiple-sense interpretation of recognition failure must be written off as a historical curiosity, an intuitively attractive idea that is not true in nature. In the latter case, the semantic interpretation becomes void of any empirical content and the theory a tautology: If *all* words and *all* names have multiple senses, the claim that recognition failure occurs for words with multiple senses is a claim that recognition failure occurs for words. The former case is accommodated by the Flexser-Tulving model (1978), the latter is consonant with our results.

We think that the multi-node generation/recognition theory is wrong, although it would be also possible to claim merely that it has limited applicability. Muter recently concurred in our interpretation, stating that our results provide "the only reasonably convincing evidence I know of that g-r theory (generation/recognition theory) is wrong" (Muter, personal communication, October 1986).

Generality of Data or a "Law" of Nature?

We have suggested that the Tulving-Wiseman function resembles an empirical "law" and have motivated our research partly in terms of the desire to seek the limits to the "law." By an empirical law we mean the same thing that other scientists do. Thus, an empirical law "seems to summarize simply some fairly directly observed regularity, without attempting to provide a theoretical explanation for it" (Holton & Brush, 1973, p. 158). Like other laws, empirical laws are formulated in terms of quantitative concepts that can be expressed in mathematical symbols. It is also conventional to expect that the conditions under which the law is supposed to hold be specified. In considering the case of the law in question, we see no reason to deviate from these time-honored precepts.

A regularity in behavioral data of the sort that we are dealing with differs in many ways from a well-established law in the physical sciences. For instance, there is much more variability. The "law" describes the behavior of the mythical "average" subject, yet in reality all subjects, and all samples of subjects, are different. Some fluctuation in observations must be attributed to these individual differences. Other fluctuations may occur because of uncontrolled variability in the conditions of the experiments. These problems are well known and need no discussion. The important point is that a remarkable degree of uniformity in the data has occurred in many experiments, despite all these potential sources of experimental error. The occurrence of such uniformity is far too rare in the psychology of memory to dismiss it as just another curiosity. Despite a plethora of regularities in all sorts of psychological data, it would be a genuine challenge to anyone to formulate a simple, quantitatively expressed statement of a relatively uniform and invariant relation between two theoretically interesting variables under specified conditions. We have used the tentative appellation of "law" in

presenting our research to the public, not only because we wish to stimulate other workers to tell us and the rest of the world why the regularity in question should not be called a "law," but also because we would like to see other proposals for "laws" and general "principles" (Cohen, 1985). We do so because we believe that this part of our science is the most neglected, most difficult, most challenging, and, in the long run, the most important (cf. Tulving, 1979).

References

Anderson, J. R. (1977). Memory for information about individuals. *Memory & Cognition, 5,* 430–442.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123.

Baddeley, A. D. (1979). Levels of processing and levels of explanation: Discussion of the papers by Treisman and Bransford, Franks, Morris and Stein. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 355–359). Hillsdale, NJ: Erlbaum.

Bartling, C. A., & Thompson, C. P. (1977). Encoding specificity: Retrieval asymmetry in the recognition failure paradigm. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 690–700.

Begg, I. (1979). Trace loss and the recognition failure of unrecalled words. *Memory & Cognition, 7,* 113–23.

Best, J. B. (1986). *Cognitive psychology.* St. Paul, MN: West Publishing Company.

Bowyer, P. A., & Humphreys, M. S. (1979). Effect of a recognition test on a subsequent recall test. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 348–359.

Brown, J. (Ed.). (1976). *Recall and recognition.* London: John Wiley & Sons.

Cohen, R. L. (1985). On the generality of the laws of memory. In L.-G. Nilsson & T. Archer (Eds.), *Perspectives on learning and memory* (pp. 247–277). Hillsdale, NJ: Erlbaum.

Donnelly, R. (1988). Priming effects in successive episodic tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 256–265.

Flexser, A. J. (1981). Homogenizing the 2 × 2 contingency table: A method for removing dependencies due to subject and item differences. *Psychological Review, 88,* 327–339.

Flexser, A. J., & Tulving, E. (1978). Retrieval independence in recognition and recall. *Psychological Review, 85,* 153–171.

Flexser, A. J., & Tulving, E. (1982). Priming and recognition failure. *Journal of Verbal Learning and Verbal Behavior, 21,* 237–248.

Gardiner, J. M., & Tulving, E. (1980). Exceptions to recognition failure of recallable words. *Journal of Verbal Learning and Verbal Behavior, 19,* 194–209.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Holton, G., & Brush, S. G. (1973). *Concepts and theories in physical science,* (Second Edition). Reading, MA: Addison-Wesley.

Horton, D. L., & Mills, C. B. (1984). Human learning and memory. *Annual Review of Psychology, 35,* 361–394.

Humphreys, M. S., & Bowyer, P. A. (1980). Sequential testing effects and the relationship between recognition and recognition failure. *Memory & Cognition, 8,* 271–277.

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing and human memory* (pp. 429–446). Hillsdale, NJ: Erlbaum.

Jones, G. V. (1983). Structure of the recall process. *Philosophical Transactions of the Royal Society, London, B, 302,* 373–385.

Jones, G. V. (1984). Analyzing recognition and recall. *Behavioral and Brain Sciences, 7,* 242–243.

Mandler, G. (1979). Organization and repetition: Organizational principles with special reference to rote learning. In L.-G. Nilsson (Ed.), *Perspectives on memory research: Essays in honor of Uppsala University's 500th Anniversary* (pp. 293–327). Hillsdale, NJ: Erlbaum.

Martin, E. (1975). Generation-recognition theory and the encoding specificity principle. *Psychological Review, 82,* 150–153.

McCormack, P. D. (1972). Recognition memory: How complex a retrieval system? *Canadian Journal of Psychology, 26,* 19–41.

Murdock, B. B., Jr. (1974). *Human memory: Theory and data.* Potomac, MD: Erlbaum.

Muter, P. (1978). Recognition failure of recallable words in semantic memory. *Memory & Cognition, 6,* 9–12.

Muter, P. (1984). Recognition and recall of words with a single meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 198–202.

Neely, J. H., & Payne, D. G. (1983). A direct comparison of recognition failure rates for recallable names in episodic and semantic memory tests. *Memory & Cognition, 11,* 161–171.

Nilsson, L.-G., Dinniwell, M., & Tulving, E. (1987). Recognition failure of categorized words. *Memory & Cognition, 15,* 389–396.

Nilsson, L.-G., & Shaps, L. P. (1980). A functional view of memory. In F. Klix & J. Hoffman (Eds.), *Cognition and memory: Interdisciplinary research of human memory activities* (pp. 40–46). Berlin: Deutscher Verlag der Wissenschaften.

Nilsson, L.-G., & Shaps, L. P. (1981). A reconstructive processing interpretation of the recognition failure phenomenon. *Acta Psychologica, 47,* 25–37.

Nilsson, L.-G., & Tulving, E. (1986, November). Exceptions to recognition failure. Paper presented at the Psychonomic Society Meeting, New Orleans, LA.

Postman, L. (1975). Tests of the generality of the principle of encoding specificity. *Memory & Cognition, 3,* 663–672.

Rabinowitz, J. C., (1984). Aging and recognition failure. *Journal of Gerontology, 39,* 65–71.

Rabinowitz, J. C., Mandler, G. & Barsalou, L. W. (1977). Recognition failure: Another case of retrieval failure. *Journal of Verbal Learning and Verbal Behavior, 16,* 639–663.

Reder, L. M., Anderson, J. R., & Bjork, R. A. (1974). A semantic interpretation of encoding specificity. *Journal of Experimental Psychology, 102,* 648–656.

Stern, L. (1985). *The structures and strategies of human memory.* Homewood, IL: The Dorsey Press.

Tulving, E. (1979). Memory research: What kind of progress? In L.-G. Nilsson (Ed.), *Perspectives on memory research: Essays in honor of Uppsala University's 500th anniversary* (pp. 19–34). Hillsdale, NJ: Erlbaum.

Tulving, (1983). *Elements of episodic memory.* New York: Oxford University Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

Tulving, E., & Watkins, O. C. (1977). Recognition failure of words with a single meaning. *Memory & Cognition, 5,* 513–522.

Tulving, E., & Wiseman, S. (1975). Relation between recognition and recognition failure of recallable words. *Bulletin of the Psychonomic Society, 6,* 79–82.

Wallace, W. P. (1978). Recognition failure of recallable words and recognizable words. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 441–52.

Watkins, M. J. & Tulving, E. (1975). Episodic memory: When recognition fails. *Journal of Experimental Psychology: General, 104,* 5–29.

Wiseman, S., & Tulving, E. (1975). A test of confusion theory of encoding specificity. *Journal of Verbal Learning and Verbal Behavior, 14,* 370–381.

Wiseman, S., & Tulving, E. (1976). Encoding specificity: Relation between recall superiority and recognition failure. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 349–361.

Inconsistencies and Aggravations: Word Processing Manuscripts in APA Style

If some of the requirements of APA editorial style and typing format, as described in the *Publication Manual,* conflict with the capabilities of your word processing system, the APA Journals Office would like to hear from you. We are reexamining style and format requirements in the the light of how easy or how difficult they are to implement by those preparing manuscripts on computers or word processors. For example, some systems cannot easily place the page number under the short title in the upper-right corner of each page. If you have had difficulties with this requirement or any others, we would much appreciate your taking the time to jot down which requirements have presented problems, and what hardware and software systems you used to prepare the manuscript. Write to Leslie Cameron, Room 710, APA, 1400 North Uhle Street, Arlington, Virginia 22201.