Contingent Dissociation Between Recognition and Fragment Completion: The Method of Triangulation

C. A. Gordon Hayman and Endel Tulving
University of Toronto, Toronto, Ontario, Canada

Two experiments conforming to the logic of the method of triangulation were conducted. Following the study of a list of words, the first of two successive tests (recognition) was identical for two groups of subjects, but the second one, in which the same word-fragment cues were presented to both groups, differed with respect to retrieval instructions. Subjects in one group engaged in cued recall of study-list words, whereas those in the second group completed the fragments with the first word that came to mind. Both experiments yielded the same result: The dependency between the first and second tests, indexed by Yule's Q statistic, was greater for recognition and cued recall than it was for recognition and fragment completion. These results speak to the controversial issue of the usefulness of contingency analyses of data from successive memory tests. The results are interpreted in a theoretical framework consisting of an integration of the idea of a hypothetical quasi-memory system with the transfer-appropriate procedural approach.

A continuing problem in research on direct priming and its effects concerns the relation between priming and episodic memory. Existing literature reveals both similarities and differences in the ways in which measures of priming and measures of episodic memory vary as a function of experimental conditions (Richardson-Klavehn & Bjork, 1988; Schacter, 1987; Shimamura, 1986). In their quest for understanding of priming and its phenomena, students of memory have naturally concentrated on the differences between priming and episodic memory rather than on their similarities, for a simple reason. Similarities between measures of memory performance hardly ever help to distinguish between theories, whereas differences may.

One particular kind of difference is revealed by analyzing the data from experiments conforming to the successive-testing paradigm. In a typical experiment, after the presentation of a list of words for study, subjects are given an episodic yes–no recognition test, followed by a second test in which priming effects are measured. Performance on the second test can be compared for those target words that were correctly recognized on the first test and those that were not. Such contingency analyses from a number of experiments have shown that the second test performance differs little for recognized and nonrecognized target words, that is, that recognition and the second test performance are stochastically independent. A variety of second tests have been used in experiments yielding such stochastic independence: word-fragment completion (e.g., Chandler, 1983, summarized in

Tulving, 1985a; also, see Light, Singh, & Capps, 1986; Tulving, Schacter, & Stark, 1982), tachistoscopic identification of words (Jacoby & Witherspoon, 1982), spelling of auditorily perceived homophones as target words (Eich, 1984), and speed of picture naming (Mitchell & Brown, 1988). Similar contingency analyses between successive episodic-memory tests almost invariably reveal positive association or dependency (e.g., Ogilvie, Tulving, Paskowitz, & Jones, 1980; Postman, Jenkins, & Postman, 1948; Rabinowitz, Mandler, & Patterson, 1977; Tulving & Watkins, 1975; Wallace, 1978; Watkins & Todres, 1978; Watkins & Tulving, 1975).[1] The juxtaposition of independence between episodic recognition and presumably nonepisodic "priming tasks," on the one hand, and the dependence between measures of episodic memory, on the other, suggests that processes underlying priming differ from those underlying retrieval of episodic information in theoretically interesting ways. It also underscores the usefulness of the method of successive testing and the attendant contingency analyses.

Possible Problems With Contingency Analyses

Two possible problems with contingency analyses have been discussed in the literature. One has to do with the so-called Simpson's paradox, the other with the so-called test-priming effects.

Simpson's Paradox

The main proponent of the idea that the existence of Simpson's paradox places serious constraints on the interpre-

---

[1] An interesting exception of no association between recognition and free recall, the reasons for which are not yet clear, has been reported by Broadbent and Broadbent (1975). For discussion, see Rabinowitz, Mandler, and Patterson (1977), Broadbent and Broadbent (1977), Watkins and Todres (1978), and Gardiner (1988).

tation of the data from successive-testing paradigms is Hintz-man (1978, 1980). Simpson's paradox refers to the mathematical fact that in some situations the dependence between two variables calculated for a particular sample of data need not faithfully reflect the dependence among subsets of the same sample. For example, when the values of the two variables change systematically across the subsets, it is possible to find two variables X and Y to be associated in the whole sample, but to be independent in subsets of the whole sample. Given such a state of affairs, Hintzman argued, contingency data are largely uninterpretable, and if "done at all, they should be interpreted with extreme caution" (Hintzman, 1980, p. 410).

Although Hintzman's arguments are technically correct, he has pointed out only an immanent possibility, something that might be. Hintzman has not produced evidence that links the possibility to reality in the kind of experiment against which he protested. Because of his failure to back up his hypothetical claims with actual facts, his overall message has been labeled as one of "hopelessness" (Martin, 1981, p. 373). When attempts have been made by others to test for the presence of Simpson's paradox in real data (e.g., Ogilvie et al., 1980), no such presence has been discerned. In our laboratory, we have performed many contingency analyses of data from successive-testing experiments, going back all the way to Tulving and Thomson (1973), and we have never found that subset analyses yield data leading to conclusions different from those based on the whole-sample data. This is why we are skeptical about the psychological relevance of Hintzman's (1978, 1980) logical points. Furthermore, a mathematical method exists (Flexser, 1981) for adjusting the data for subject and item differences, which, according to Hintzman, are major sources of the hidden covariates responsible for Simpson's paradox.

Despite the conjectural nature of Hintzman's (1978, 1980) criticism of contingency analyses, and despite the lack of any evidence to support it, his ideas seem to have struck responsive chords in the souls of some investigators. We know of several instances in which manuscripts submitted for publication have been rejected on the advice of anonymous reviewers who, citing Hintzman (1980), argue that contingency analyses of data from successive memory tests are scientifically illegitimate. The only explanation we have been able to invent to account for these attitudes is that many experimental psychologists have been trained to be highly suspicious of all correlational data. This deep-seated aversion goes back through generations. From time to time it is expressed explicitly: "The correlational method is the invention of the devil" (Hilgard, 1955, p. 228).

Contaminating Effects of Test Priming

The second perceived difficulty with contingency analyses lies in so-called test-priming effects, referring to the well-established fact that the first test frequently changes the outcome of the second test (e.g., Humphreys & Bowyer, 1980; Postman, 1975; Tulving et al., 1982; Tulving & Watkins, 1975; Watkins & Todres, 1978; Watkins & Tulving, 1975). This fact suggests that the outcome of the second test provides a measure of the combined effects of the earlier study episode

and the first test. The contingency analysis of the data from the successive-testing paradigm, therefore, reflects the relation between the pure study effects of the first test and the biased or contaminated effects of the second test.

Shimamura (1985) has argued that correlations observed in the contingency analysis may reflect the effects of test priming instead of the nature of the true relations between the variables indexed by the outcomes of the two tests. Shimamura directed his comments primarily at the conclusion drawn by Tulving et al. (1982) that recognition and primed fragment completion, administered in this order, are stochastically independent. He argued that the results obtained by Tulving et al. (1982) are equivocal and invalid because they could have been obtained for either of two reasons: independence between the two measures or dependence between them that was masked by massive test priming present in the Tulving et al. (1982) experiment. Using hypothetical data and an all-or-none model of transfer, Shimamura (1985) found that if the "real" relation between recognition and primed fragment completion was one of dependence, then the effect of test priming would be to reduce the dependence observed. He suggested that such reduction may have occurred in the Tulving et al. (1982) experiment to the extent that it led the authors to the (erroneous) conclusion of independence.

Despite the fact that Shimamura (1985), like Hintzman (1978, 1980) before him, did not present any empirical evidence in support of his arguments, he and his colleagues have kept repeating the point (e.g., Shimamura, 1986; Squire, Shimamura, & Graf, 1987), and his ideas have been accepted by others. For instance, Mandler, Graf, and Kraft (1986), in trying to cope with the Tulving et al. (1982) data from the point of view of their own theory, assumed that in an experiment such as Tulving et al.'s (1982), the recognition test adds a certain "amount of activation" to the target words and that this added activation "will increase the likelihood of finding stochastic independence when in fact no such independence exists" (p. 659, italics added). In discussing these criticisms of theoretical interpretations of observed stochastic independence, Richardson-Klavehn and Bjork (1988), in their trenchant review of direct and indirect measures of memory, agreed that the problem of "contaminating effects" of the first test is "fundamental and formidable" (p. 497).

This article reports a modest effort to do something about this "fundamental and formidable" problem. We believe that the successive-testing method and contingency analyses represent useful adjuncts to other methods experimental psychologists have adopted for the study of memory, if for no other reason than that it provides a unique perspective on the operations of the mind, a perspective not afforded by any other method.[2]

---

[2] The editor of this journal, skeptical about the use of successive tests and contingency analyses, in reviewing the manuscript said that "most researchers interested in differences between implicit and explicit memory prefer functional dissociations in which one manipulates an independent variable and sees differential effects on two tests." And he wondered whether there is "something special that we can learn from successive testing that cannot be better learned some

Purpose of the Study

Our purpose in this article is to address the problem of interpretation of the data from experiments such as Tulving et al.'s (1982) in which an episodic recognition test is followed by a primed word-completion test and in which contingency analyses are performed on the outcomes of the two tests. We pursue this aim by doing four things:

1. We offer a partial solution to the problem of test priming, that is, the contaminating effects of the first test on the joint outcome of both tests, in the form of what we refer to as the "method of triangulation." The problem of test priming is minimized in this method by virtue of holding the effects of the first test constant between experimental conditions of interest.

2. We propose a reorientation of the handling of data from successive-testing paradigms: Instead of describing the outcome of contingency analysis in terms of a simple dichotomy, "dependent" or "independent" for a single association, we suggest that the outcome should be described in terms of differences in the degree of association observed in a control and an experimental comparison. We suggest that Yule's Q, a venerable measure of association in fourfold contingency tables, be adopted for the purpose of these comparisons. We remind the readers what Q is and employ it in analyzing the data from our experiments.

3. We report two experiments patterned after the Tulving et al. (1982) study in which a studied list was followed by recognition and word-completion tests. In the latter test, subjects were provided with word-fragment cues for the production of the study-list words. The critical experimental manipulation concerned instructions given to the subjects in the word-completion tests. There were two kinds of instruction, defining two different retrieval tasks, as in the third experiment reported by Graf and Mandler (1984). In the cued-recall task, subjects had to respond to the cues with words that had appeared in the study list; in the word-completion task, they could respond to the cues with any word that fit the given fragment. Thus, the cued-recall task taps episodic memory, whereas the word-completion task may tap some other memory system instead of, or in addition to,

episodic memory. The experiments yielded the same empirical result: The contingency relation between recognition and primed word-completion performance, indexed by Q, varied systematically with the task given to subjects on the second test, with other conditions of the experiment, including the retrieval cues, held constant.

4. We evaluate the objections to the successive-testing paradigm and the disagreements in the interpretation of the data yielded by contingency analyses in the light of our findings and those from other relevant experiments. We conclude that the method of triangulation constitutes a useful adjunct to other techniques that have been employed in the study of priming phenomena and the relation of these phenomena to episodic memory.

Contingent Dissociation and the Method of Triangulation

Although the two experiments reported here vary in a number of details, some important, some less so, they share certain critical features with the Tulving et al. (1982) experiment, namely two successive tests entailing different tasks for the subject and an item-by-item contingency analysis of the relation between the outcomes of the two tests. The major difference in the current experiments had to do with the instructions given to subjects at the time of the second test—cued recall versus word completion—while holding the physical form of the retrieval cues constant (cf. Graf & Mandler, 1984; Graf, Squire, & Mandler, 1984; Neely & Payne, 1983). The retrieval cues were graphemic word fragments, that is, words with a number of letters deleted from various positions in a word (e.g., Tulving et al., 1982).

Other researchers have already demonstrated that cued-recall versus word-completion instructions produce *functional dissociations*: An independent variable X may affect cued recall but not word completion (Graf & Mandler, 1984; Graf & Schacter, 1985; Schacter & Graf, 1986). Our experiments demonstrate a comparable *contingent dissociation* in what we will refer to as the *triangulation paradigm*. In this paradigm, the contingency relation is measured between a common reference test and each of two or more variable tests under otherwise fixed experimental conditions.[3] Contingent dissociation between Variable Tests A and B is demonstrated when the measured dependence between the reference test and Test A is different from the measured dependence between the reference test and Test B.

The triangulation paradigm has general applicability. Its logic holds, in principle, with any set of three tests as well as in situations in which the first (reference) test leads to differential priming of individual items. In our experiments, the reference test was yes–no recognition, and the two variable tests were cued recall and word completion. As all other experimental conditions were held constant in the variable

---

other way?" These comments seem to imply that (a) contingency analyses are *alternatives* to functional analyses of the relation between variables and (b) the data obtained with one technique are interchangeable with those obtained with the other. If these implications were true, then of course it would be more rational to "simply stick with studies of functional independence," as our esteemed colleague put it. But the assumptions are wrong. The two methods are complementary. The results obtained with one method are in theory independent of the results obtained with the other, and hence what one method yields cannot be predicted from the knowledge of what the other method has yielded. Functional independence of Tests X and Y may obtain in situations in which the contingency relation between Tests X and Y is positive, 0, or negative. Functional dependence (including perfectly parallel effects of some independent variable on Tests X and Y) can similarly obtain in situations in which the contingency relation is positive, 0, or negative. Given these facts, it is certainly possible to "stick with one method," but only if one simply does not care to know what the other method might reveal. These matters have been briefly mentioned before (Tulving, 1983, p. 104–105; Tulving, 1985a, p. 393–395).

---

[3] We have adopted *triangulation* as the designation of the paradigm because (a) we need a distinctive short label to refer to the method whose full description is long, and (b) the experimental operations in the paradigm are not unlike those involved in measuring the distance between two points in space by referring certain measurable properties of each of them (distance, angle) to a common reference point.

tests, and only retrieval instructions were varied, we could assume that any difference between the experimental conditions would reflect the consequences of different cognitive processes involved in cued recall and (primed) word completion.

Yule's Q as an Index of Dependence

The purpose of the method of triangulation is to compare two or more contingency relations. Doing so only in terms of a 2-point scale—*dependent* versus *independent*—would amount to deliberately ignoring variations in the degree or amount of dependence, variations that may be significant theoretically. To take full advantage of the method of triangulation, one must adopt an index of dependence.

A number of measures of dependence in 2 × 2 tables have been used by investigators over the years. Perhaps the most popular has been the use of conditionals and chi-square (Graf & Mandler, 1984; Graf & Schacter, 1985; Greeno, James, & DaPolito, 1971; Jacoby & Witherspoon, 1982; Martin, 1971), although some investigators have also used the phi correlation. Different measures focus on (slightly) different aspects of association, and averaged across situations, there is no single aspect that predominates (for a detailed discussion of these issues, see Bishop, Fienberg, & Holland, 1975; Fienberg, 1977; Nelson, 1984; Upton, 1978).

We think that Yule's Q (Yule, 1900) is preferable to other measures because it avoids some problems inherent in other measures. The problem with chi-square is that it is a measure of statistical significance and, as such, varies as a function of the number of events sampled. The major problem with phi is that it confounds the marginal relation between categorical variables with the interaction between these variables (see Fienberg, 1977; Nelson, 1984; Upton, 1978). That is, phi can express strong positive dependence only when the proportions correct on the first and second test are equal or close to equal, and it can express the full -1 to $+1$ range of dependence only when both probabilities are equal to .50. The Q statistic has neither of these shortcomings.

The Q statistic, the version of gamma (Goodman & Kruskal, 1954; see also Nelson, 1984) applied to the data from fourfold contingency tables, reflects the difference in the conditional probabilities of observing concordant and discordant responses in a 2 × 2 table. If the joint event that both responses are successes is represented by a, the joint event that both responses are failures by d, and the two joint events that one response is a success and the other a failure by b and c, then Q may be calculated by the following formula:

$$\text{Yule's } Q = (ad - bc)/(ad + bc). \tag{1}$$

For comparison, phi or Pearson's r correlation computed on a 2 × 2 table would be written as:

$$\text{Phi} = (ad - bc)/[(a + b)(c + d)(a + c)(b + d)]^{1/2}. \tag{2}$$

As can be seen, the numerator or estimate of covariance ($ad - bc$) is the same for both measures of association, but they differ in terms of the denominator or the estimate of maximum variance.

The range of Q (-1 to 1), like phi, is symmetric about 0, with the end points corresponding to complete (negative or positive) association and with 0 corresponding to the condition of no association. However, unlike phi, the range of Q remains invariant for all combinations of marginal values.

Although a method of calculating the variance and, hence, the significance, of Q is given by Upton (1978, p. 18), the distribution of Q converges to normal slowly and requires increasingly larger samples to converge to normal as Q increases. For this reason, in the present article we will report the standard deviation of Q, as given in Equation 3, but use a more conservative test of the significance of Q, namely a chi-square test statistic based on a related statistic, the *log odds ratio*, given in Equation 4.

$$SD(Q) = [1/4(1 - Q^2)^2(1/a + 1/b + 1/c + 1/d)]^{1/2}. \tag{3}$$

$$\chi^2 = [\ln(ad/bc)]^2/(1/a + 1/b + 1/c + 1/d). \tag{4}$$

The *odds ratio*, or *cross-product ratio*, is the maximum likelihood estimate of an interaction (independent of the marginal values) between two attributes in a 2 × 2 table, for multinomial, product-multinomial, and Poisson sampling models (Fienberg, 1977; Upton, 1978). However, dependence as represented by the odds ratio is nonsymmetric about the point of independence. It is for this reason of scaling that we chose Q, a monotonic function of the odds ratio, as our measure of dependence rather than the odds ratio itself.

In addition to assessing dependence or independence within a table, in the research reported here, we will find it necessary to compare the degree of dependence between two 2 × 2 tables. The method we have adopted for this purpose consists of testing for a second-order interaction using the log-linear model, as discussed by Bishop et al. (1975), Fienberg (1977), and Upton (1978). The computational formula for the test statistic of this interaction is given in the following equation:

$$\chi^2 = (\log C_1 - \log C_2)^2/[V(C_1) + V(C_2)], \tag{5}$$

where C_1 is the odds ratio for a matrix given by Equation 6, and $V(C_1)$ is the large sample variance of the natural logarithm of C_1 given by Equation 7.[4]

$$C = (ad/bc) = \text{odds ratio.} \tag{6}$$

$$V(C) = (1/a + 1/b + 1/c + 1/d) = \text{variance of } C. \tag{7}$$

If two odds ratios are equal, then the test statistic from Equation 5 will have an asymptotic chi-square distribution with 1 degree of freedom. In the experiments that follow, Equation 5 is used to provide a test for a second-order

---

[4] A concern expressed by one of the reviewers was that the chi-square test comparing differences between Yule's Q seemed to violate the elementary independence assumption. The assumption of independent observations is critical only in identifying the source of an observed nonzero dependence, not in estimating the likelihood of that event. That is, a significant chi-square is a reliable indicator of a difference in association, regardless of the source(s) of the dependence. And it is insensitive to whether the difference in association between two nonzero associations or between a nonzero association and a zero association can be further decomposed into more than one source of dependence.

interaction in the degree of association between two 2 × 2 tables. Values less than 3.84 will be taken as representing insufficient evidence for rejecting the null hypothesis that the levels of association in two 2 × 2 tables are sampled from the same distribution. Values greater than 3.84 will be taken as evidence that levels of association represented in the two tables are derived from different distributions.

In summary, we have adopted Q as an index of association because it has several advantages over other possible measures: (a) Its range is bounded by -1 and $+1$, (b) tests of significance of Q exist, and (c) tests of differences of Q exist. The latter feature is particularly important for our present purposes: Given two experimental conditions in which Q is obtained, we wish to know how reliable observed differences in Qs are.

Overview of Experiments

The experiments were designed to compare the contingency relation between recognition and fragment-cued recall with the contingency relation between recognition and fragment-cued word completion. In both experiments, the study list was presented to the subjects on a single trial. The priming effects observed from the study presentation, therefore, reflect the consequences of a single brief encounter with the target word. In both experiments, the critical items were familiar English words. Materials were always counterbalanced across the experimental conditions, with the consequence that no differences in measured performance can be attributed to differences in the tested words. In other respects, too, appropriate controls were included in the designs of the experiments to eliminate unwanted confoundings. Within any given experiment, retrieval cues provided in the cued-recall and primed word-completion tests were the same, and only the instructions to the subjects as to what they had to do with the cues varied. The scoring of the subjects' performance in the two retrieval tasks was also identical: Target words for which subjects were given credit were the words that had appeared on the study list. Our general plan for experiments conforms to the methodological suggestions made by Neely and Payne (1983) and follows the example set by Graf and Mandler (1984) and Graf et al. (1984).

We will report the main data from our experiments in sufficient detail for the reader to be able to follow the connection between the initial tabulation of the data in the 2 × 2 contingency tables and the conclusions we will draw.

Experiment 1

The main purpose of Experiment 1 was to demonstrate the applicability of the triangulation paradigm to a comparison of cued recall and primed completion. In the design, two groups of subjects would study the same list and would take the same recognition test. Then one group would take the cued-recall test, whereas the other would take the completion test. The actual retrieval cues would be the same in both groups (word fragments), and only the retrieval instructions would vary. These instructions presumably affect what information is retrieved, how it is retrieved, or both (cf. Tulving,

1983, p. 175–176). From a methodological point of view, the finding of differential dependency could not be dismissed as merely a consequence of successive tests or contingency analysis. From a theoretical point of view, the observed differential dependency would become yet another touchstone against which competing conceptualizations of human memory can be compared.

A secondary purpose of the experiment was to examine whether the expected differential dependency would hold for part-word cues that permitted more than a single completion. To this end, two types of fragment were employed: fragments with relatively few solutions and fragments with a large number of solutions.

Method

Subjects. Thirty-two introductory psychology students at the University of Toronto participated in the experiment for course credit. Subjects were tested individually and were assigned randomly to one of two groups.

Materials. The stimuli consisted of 128 pairs of four-letter words that differed by only one letter, taken with a few modifications from Johnston (1978, Experiment 1). Examples are *drip, grip* and *date, gate.* The three letters common to a given pair formed a three-letter fragment cue, identical for both members of a pair. One member of each word pair was arbitrarily designated as the *target* word and the other member as the *alternate* word. For counterbalancing purposes, these 128 pairs were randomly divided into four subsets of 32 pairs.

There were two main types of word fragments: (a) high-constraint fragments with an average of 2.7 possible word completions and (b) low-constraint fragments with an average of 6.4 possible word completions. The word pairs and fragments were selected with three basic constraints: (a) None of the target words could serve as a completion for more than one fragment; (b) for each high-constraint fragment and word pair (e.g., *rip, drip,* and *grip*) there was a matching low-constraint fragment and word pair with the missing letter in the same position (e.g., *ate, gate,* and *date*); and (c) letters from each of the four letter positions were deleted equally often in constructing the fragments. An additional set of 32 four-letter words was selected from the same source to serve as buffer items in the study session, with the constraint that no buffer word could serve as a completion for a target fragment.

Design and procedure. The design was a 2 × 3 mixed factorial, with one factor between and one factor within subjects. The between-subjects manipulation was the type of instructions given for the fragment-cued responses (cued recall or fragment completion), and the within-subjects manipulation was presentation history (target word tested in recognition, alternate word tested in recognition, and neither target nor alternate word tested in recognition). Each subject studied a list of 128 words (16 primacy buffers, 96 target words, and 16 recency buffers) presented at the rate of one word every 2 s on a CRT monitor controlled by an APPLE II/e computer. The target words consisted of a random ordering of all the words from three subsets.

Following the study list, subjects were engaged in casual conversation. Three minutes after the last study word was presented, all subjects were given the first test of retention, a 96-word yes–no recognition test, composed of 32 studied (old) and 64 experimentally novel (new) words. Two types of new words were presented in the recognition test: (a) 32 alternate words whose paired target word was presented during study and 32 words from the fourth subset whose target word was not presented during study.

Three minutes after the last recognition judgment, the second test commenced. For both types of second test, subjects were presented with 96 three-letter fragments for a maximum of 8 s per fragment. All part-word cues could be completed by a study word, for both subject conditions. Subjects in the fragment-cued recall group were told of this arrangement and were instructed to use the fragment as a cue to recall one of the study (target) words. Subjects in the fragment-completion group were informed that although many of the solution words for the fragments had been presented during study, many other solutions had not been seen during the course of the experiment. They were instructed to complete the fragment by naming any English word that came to mind. Otherwise, subjects in the two groups received identical treatment and stimuli. The tested fragments fell into one of three conditions: (a) target at study and recognition, (b) target at study and alternate at recognition, and (c) target at study. Fragment-cued responses were analyzed separately for target and alternate words by analyzing the proportion correct in $2 \times 2 \times 3$ mixed-plot analyses of variance (ANOVAs; Instructions × Constraint × Presentation History).

Results and Discussion

Because instructions were manipulated after the recognition test, there were no significant differences in recognition hits (Ms = .789 and .755 for recall and completion instructions, respectively) or false alarms (Ms = .352 and .273 for recall and completion instructions), both Fs < 1. In both experiments, all results are reported at the $p < .05$ level of significance unless otherwise indicated.

The proportion of target and alternate words produced by the two instruction groups, for the two types of fragments, and for the three test conditions are presented in Table 1.

Overall, there were two major influences on target-word responses to fragment cues. Responses were more frequent for high- (M = .461) than for low-constraint fragments (M = .326), $F(1, 30) = 90.56$, $MS_c = .010$, and responses were more frequent if the target word was presented both at study and at recognition, $F(2, 60) = 94.96$, $MS_c = .011$, Fisher's LSD = .075.

Although there was no main effect of, or interaction involving, instructions when all three target conditions were compared, a separate analysis of target words presented both at study and on the recognition test showed that cued recall (M = .584) was greater than completion (M = .490), $F(1, 30)$

Table 1

Mean Proportion of Target- and Alternate-Word Responses to Fragment Cues in Experiment 1

Presentation	ST–RT		ST–RA		ST	
	HiC	LoC	HiC	LoC	HiC	LoC
Cued-recall instructions						
Target	.633	.535	.352	.254	.441	.289
Alternate	.090	.094	.328	.316	.144	.133
Completion instructions						
Target	.563	.418	.340	.196	.438	.262
Alternate	.238	.168	.484	.348	.340	.211

Note. HiC = high constraint on fragment solution; LoC = low constraint on fragment solution; ST = studied target word; RT = recognition tested with target word; RA = recognition tested with alternate word.

= 5.47, MS_c = .026. The interaction with fragment type (high constraint vs. low constraint) was not significant, F < 1. These results replicate, with word fragments, the observation by Graf et al. (1984) that cued recall using word stems as cues can be higher than primed stem completion.

There were two main influences on the production of alternate words. First, there was a significant effect of instructions, $F(1, 30) = 40.26$, $MS_c = .015$, where subjects with recall instructions produced fewer alternate words (M = .184) than did subjects with completion instructions (M = .298). Thus, to some degree, subjects in the recall group obeyed the instructions to produce only study words. Second, and not surprisingly, alternate words presented as recognition lures were produced more frequently, $F(2, 60) = 75.56$, $MS_c = .011$, Fisher's LSD = .076. In addition, there was an unexpected interaction between instructions and constraint, $F(1, 30) = 16.48$, $MS_c = .008$, Fisher's LSD = .065, where, perhaps reflecting floor effects, there was no difference between high- and low-constraint fragments in cued recall (Ms = .187 and .181, respectively) but a significant difference in completion (Ms = .354 and .242 for high- and low-constraint fragments).

Dependence. The relevant data are presented in Table 2. When subjects were instructed to use the fragment as an aid for recall, there was a statistically significant dependence between recognition and fragment-cued responses (Qs = .553 and .377 for high- and low-constraint fragments, respectively). When subjects were instructed to complete the fragment with any English word, there was no systematic relation between recognition and fragment-cued responses (Qs = .182 and −.167 for high- and low-constraint fragments). These differences in dependence as a function of instructions were significant, $\chi^2 (1, N = 256) = 3.92$ and 7.17 for high- and low-constraint fragments. In short, fragment-completion responses were essentially independent of prior recognition responses, whereas fragment-cued recall responses were associated with prior recognition responses.

Experiment 2

Experiment 2 was designed as a replication of Experiment 1, with different materials. The design was the same as that used in Experiment 1, except that the study-list words and the cues in the second test were of the kind used by Tulving et al. (1982). The degree of constraint on completion was manipulated by using fragments allowing only one completion and fragments allowing two completions.

Method

Subjects. Thirty-two introductory psychology students at the University of Toronto participated in the experiment for course credit. Subjects were tested individually and assigned randomly to one of two groups.

Materials. The stimuli consisted of 64 words and corresponding fragments with a single-word completion and 64 words and corresponding fragments with two-word completions. The single-word fragments (high constraint) and associated words were a subset of those used by Tulving et al. (1982), selected with the constraint that the probability of fragment completion was neither too high nor too

234 C. A. GORDON HAYMAN AND ENDEL TULVING

Table 2
Dependence Between Recognition and Fragment-Cued Responses as a Function of Instructions in Experiment 1

Instructions	Rn	CR	RnCR	Rn*CR	Q	SD(Q)	χ^2
Cued recall							
HiC	.793	.633	.551	.502	.553	.111	15.13
LoC	.785	.535	.453	.420	.377	.134	6.45
Completion							
HiC	.785	.563	.457	.442	.182	.148	1.45
LoC	.730	.418	.289	.305	−.167	.138	1.41

Note. HiC = high constraint on fragment solution; LoC = low constraint on fragment solution; Rn = recognition; CR = fragment cued response; RnCR = recognized and recalled; RN*CR = expected recognized and recalled when independent; N for each row = 256.

low. These stimuli were seven and eight letters in length, and fragments were created by deleting three letters in seven-letter words and four letters in eight-letter words (e.g., *assassin, a--a--in*). The two-word fragments (low constraint) consisted of five- to eight-letter words with three to four letters removed (e.g., *polite, police, p--i-e*). On the basis of norms of completion,[5] two sets of 64 words were created: The target set consisted of the less likely of the two-word solutions, and the alternate set consisted of the more likely solution. Only words from the target set were presented during the course of the experiment. Fragment-completion responses were scored in terms of both the target- and the alternate-word sets.

Four subsets of 32 words were constructed, with 16 high-constraint and 16 low-constraint words in each. The lists were counterbalanced across subjects such that each list occurred an equal number of times in each of the four test conditions. An additional 128 words were used as primacy and recency buffers during presentation of the study list.

Design and procedure. The design was a 2 × 2 × 2 × 2 mixed factorial, with one between- and three within-subjects factors. Test instructions (cued recall or fragment completion) were manipulated between subjects, whereas fragment type (high or low constraint), study presentation (old or new), and recognition (tested or nontested) were manipulated within subjects.

Each subject studied a list of 192 words (32 primacy buffers, 64 target words, and 96 recency buffers). The buffer words were included to avoid ceiling effects in recognition. The target words were composed of all the words from two subsets and were presented at the rate of one word every 2 s, as were the buffer words. Subjects were engaged in casual conversation after presentation of the study list. Three minutes after presentation of the last study word, all subjects were given a yes–no recognition test that consisted of a random order of 32 old words (from one of the studied subsets) and 32 new words (items from one of the remaining two subsets). Subjects were engaged in casual conversation after finishing the first test. The second test, fragment-cued recall or fragment completion, began 5 min after the last recognition judgment. This test consisted of 128 fragments presented for a maximum duration of 8 s. The only difference between the two groups was in the instructions given prior to presentation of the fragments. Subjects in the fragment-cued recall group were informed that although many of the fragment cues could be solved in terms of two or more words, their task was to use the fragment as an aid in recalling study words. Subjects were further instructed that if they thought of a word they did not remember as having been presented in the study list, they were to ignore this word and were to attempt another solution. Subjects given fragment-completion instructions were told to complete the fragment by giving any English word that came to mind. In addition, these subjects were instructed that although some of the fragments could be completed with words

previously seen in the experiment, many of the fragments could not be solved in this manner, and their task was to solve as many of these fragments as possible. Otherwise, subjects in the two groups received identical treatment and stimuli. Both high- and low-constraint fragments were tested in one of four conditions: (a) study and recognition priming of the target words, (b) study priming of the target word, (c) recognition priming of the target word, and (d) no priming of the target word.

Study and test stimuli were presented on a CRT monitor under the control of an Apple II/e computer. Fragment-cued responses were analyzed separately for target words by analyzing the proportion correct in a 2 × 2 × 2 × 2 mixed-plot ANOVA (Instructions × Fragment Type × Study Presentation × Recognition Tested) and for alternate words by analyzing the proportion correct in a 2 × 2 × 2 mixed-plot ANOVA (Instructions × Study Presentation × Recognition Tested).

Results and Discussion

As might be expected, because the manipulation of instructions followed the recognition test, there were no significant differences in recognition attributable to instructions (Ms = .613 and .629 for recall and completion groups, respectively) or in false alarms (Ms = .258 and .207), both Fs < 1. There was, however, an interaction between study presentation (hits, false alarms) and fragment type (high and low constraint), $F(1, 30) = 9.72$, $MS_e = .013$, Fisher's LSD = .083, reflecting the larger difference between hits and false alarms for high- (M = .448) than for low-constraint fragments (M = .321). This difference may have been due to a difference in word frequency between the two sets of response words associated with the two types of fragments. Word frequency was not counterbalanced between these two types of stimuli.

Table 3 summarizes the mean proportion of fragment-cued responses for the 16 target- and 8 alternate-word conditions. There are two results of interest. First, there were no significant differences in target-word responses due to instructions, F < 1. Thus, for these materials there appears to be no major advantage of cued recall over completion, or vice versa. Second, cued-recall instructions yielded fewer alternate words (M = .218) than did completion instructions (M = .324), $F(1, 30) = 26.79$, $MS_e = 3.61$. As in the previous experiment, subjects performing the cued-recall task produced fewer non-studied words as intrusions to fragment cues. This fact suggests that recall subjects succeeded at least to some extent in following the instructions to produce study words.

Dependence. Contingency data are summarized in Table 4. As in the first experiment, there was a marked effect of instructions on the relation between recognition and fragment-cued performance. Recognition and fragment-cued recall were significantly dependent (Qs = .568 and .389, for high- and low-constraint fragments, respectively), whereas recognition and fragment completion were not (Qs = .229 and .004, for high- and low-constraint fragments). The differential dependence between recall and completion instructions was reliable for both fragment types, $\chi^2(1, N = 256) = 4.64$ and 4.68 for high- and low-constraint fragments. These results replicate the differential dependence observed in Experiment

---

[5] We would like to thank Michael Masson and Ruth Donnelly for providing the word-completion fragments with two solutions and corresponding completion norms used in Experiment 2.

Table 3

Mean Proportion of Target- and Alternate-Word Responses to Fragment Cues in Experiment 2

| | HiC fragments | | | | LoC fragments | | | |
| | Studied | | New words | | Studied | | New words | |
Instructions	Rn	NRn	Rn	NRn	Rn	NRn	Rn	NRn
Cued-recall								
Target	.535	.297	.389	.109	.418	.313	.348	.125
Alternate					.211	.176	.203	.281
Completion								
Target	.493	.324	.410	.121	.375	.312	.312	.180
Alternate					.310	.243	.356	.391

Note. HiC = high constraint on fragment solution; LoC = low constraint on fragment solution; Rn = tested in recognition; NRn = not tested in recognition.

1 with different materials and provide additional evidence that the degree of dependence between sequential tests of memory varies systematically with the type of retrieval processes in which subjects engage. With all other conditions held constant, instructions to the subjects as to what they have to do with the retrieval cue influence the nature of retrieved information, or the nature of retrieval operations employed in the second test, with consequent differences in the contingency relations between successive tests.

Other results. Overall, the pattern of results given in Table 3 parallels the results of the recognition followed by fragment completion test order in Tulving et al. (1982). There were significant main effects of study presentation and recognition testing as well as an interaction between study and recognition, $F(3, 30) = 9.37$, $MS_e = 0.14$. But there was also an interaction for target words between fragment type and recognition test, $F(3, 30) = 14.77$, $MS_e = .015$, with recognition-tested words showing a larger facilitation for high-constraint than for low-constraint fragments. Finally, there was a significant interaction between study and recognition presentation for alternate-word solutions, $F(1, 30) = 10.31$, $MS_e = .071$, indicating that alternate-word production was an inverse function of target-word production. A similar pattern was observed in Experiment 1. All other main effects and interactions resulted in nonsignificant F values.

The results of Experiment 2 replicated the differential dependency observed in Experiment 1 with two additional sets

Table 4

Dependence Between Recognition and Fragment-Cued Responses as a Function of Instructions in Experiment 2

Instructions	Rn	CR	RnCR	Rn*CR	Q	SD(Q)	χ^2
Cued-recall							
HiC	.648	.535	.418	.347	.568	.094	21.76
LoC	.578	.418	.289	.242	.389	.113	9.54
Completion							
HiC	.645	.492	.344	.317	.229	.125	3.13
LoC	.613	.375	.230	.230	.004	.133	.00

Note. HiC = high constraint on fragment solution; LoC = low constraint on fragment solution; Rn = recognition; CR = fragment cued response; RnCR = recognized and recalled; RN*CR = expected recognized and recalled when independent; N for each row = 256.

of materials. Unlike Experiment 1, however, the increased dependence observed with cued recall was not associated with higher cued recall than primed word completion. Although there was a small trend in this direction, for both single and double completion fragments, it was not statistically reliable. Thus, the instructional manipulation of differential dependence with word-fragment cues does not appear to reflect changes in processing that necessarily result in higher fragment-cued recall than primed fragment completion, nor does it appear to reflect changes in processing that necessarily result in lower fragment-cued recall than primed fragment completion.

General Discussion

The major purpose of this article was to evaluate the criticisms advanced against the use of the successive-testing paradigm and contingency analyses of data from such a paradigm, with special reference to experiments of the kind reported by Tulving et al. (1982) in which episodic recognition was found to be stochastically independent of primed fragment completion. In the pursuit of this aim, we conducted two experiments, using what we dubbed the method of triangulation, in which the contingency relations between a standard test and two or more different tests are measured and compared. We used Yule's Q as the index of dependence between the two successive tests.

The results of our experiments showed that (a) the contingency relation between recognition and cued recall was one of dependence, and (b) the degree of this dependence was higher than that between recognition and fragment completion.[6] Indeed, the dependence observed between recognition and fragment completion was not reliably different from 0.

In this section, we relate our results to the criticisms that prompted our study and briefly examine the theoretical implications of our findings. Our first finding, that recognition and cued-recall responses are dependent, is only modestly important insofar as it rules out the possibility that the format of the retrieval cue is a critical determinant of the nature of the relation between whole-word recognition and part-word recall: Using the same cue but different retrieval instructions, the dependence was greatly reduced. Many previous experiments have shown dependence between two successive episodic-memory tests, such as recognition, cued recognition, free recall, and cued recall (e.g., Darley & Murdock, 1971; Mandler, Pearlstone, & Koopmans, 1969; Ogilvie et al., 1980; Postman et al., 1948; Rabinowitz et al., 1977; Tulving & Thomson, 1973, Tulving & Watkins, 1975; Wallace, 1978; Watkins & Todres, 1978; Watkins & Tulving, 1975). It is in comparison with all these perfectly sensible and intuitively reasonable observations that the original findings of stochastic independence between recognition and primed performance (e.g., Jacoby & Witherspoon, 1982; Tulving et al., 1982) can be regarded as unexpected, interesting, and worthy of further thought. Mere observation of stochastic independence, in and of itself, is not noteworthy.

---

[6] We have conducted one other experiment similar to Experiments 1 and 2 whose findings were similar to those reported here.

It may be worth noting that the degree of dependence between recognition and cued recall in our experiments was very much in line with the Tulving–Wiseman function (Tulving & Wiseman, 1975). For the four conditions in the two experiments in which recognition was followed by cued recall (data summarized in Tables 2 and 4), the mean probability of recognition given recall was .80, exactly the same value as the mean of the same four conditional probabilities "predicted" by the function.

Such a high degree of agreement is not entirely unexpected. It is now known that the Tulving–Wiseman function describes the outcome of a large number of experiments in which a recognition test for studied words has been followed by a cued-recall test (Tulving, 1983), provided that the cue–target relation is encoded at study and provided that the retrieval information available to the subject in one test is largely independent of that in the other (Flexser & Tulving, 1978, 1982; Nilsson, Dinniwell, & Tulving, 1987; Nilsson, Law, & Tulving, 1988). Although the fragment cue in the recall test is nominally a part of the whole-word cue in the recognition test, it is not functionally so: Given one cue, it would be difficult for the subject to generate the other independently of relevant episodic information. Hence the same relation between recognition and recall was observed in our two experiments as has been observed in many recognition-failure experiments. (For further discussion, see Nilsson et al., 1988.)

The main empirical result of our experiments is the measured differential dependency.

Differential Dependency

The shorthand expression of *differential dependency* refers to the fact, observed in both of our experiments, that the dependence between recognition and fragment-cued word completion was smaller than the dependence between recognition and fragment-cued recall. Our review of the literature suggested that such differential dependency exists, but this tentative conclusion was based on post hoc, between-experiments comparisons. Under the circumstances it was possible that a variable other than the nature of the retrieval task was responsible for the outcome of the relevant comparisons.

In our two experiments, differential dependency was found under conditions in which all relevant variables were held constant, with the single exception of retrieval instructions. Therefore, it seems reasonable to conclude that the difference in dependence between recognition and cued recall, on the one hand, and recognition and fragment completion, on the other hand, is attributable to (a) differences in the kind of information tapped by the two second tests, (b) differences in processes involved in retrieval of the same kind of information, or (c) both. We will return to these issues later. Here we would like to note that the finding of differential dependency has important implications for the criticisms that have been leveled against contingency analyses in general (e.g., Hintzman, 1980) and against earlier findings of stochastic independence between recognition and fragment completion in particular (e.g., Mandler et al., 1986; Shimamura, 1985).

Consider first Hintzman's (1980) arguments that the potential for confounding from certain hidden covariates, respon-

sible for Simpson's paradox, effectively rules out any useful interpretation of contingency relations in the successive-testing paradigm and his observation that the contingency relation between the two variables or tests under scrutiny may be spuriously high or low because of the failure to take into account the effects of these covariates.

Hintzman's (1980) arguments may have potential relevance to situations in which single contingency relations are observed and interpreted, such as the finding of stochastic independence between recognition and fragment completion in the Tulving et al. (1982) experiment. But in the experiments described here, we assessed the dependence between a standard task and a variable task as the dependent variable and observed variations in it as a function of the nature of the variable task. The finding that this dependence was higher for cued recall than fragment completion as the variable task cannot be readily accounted for by invoking the possibility of Simpson's paradox. Although it is possible to explain differential dependency in terms of different hidden covariates in the two cases—cued recall and fragment completion—such an explanation simply provides a psychological explanation of the outcome of a particular contingency analysis and, in doing so, acknowledges the legitimacy of contingency analyses.

Consider next the criticisms of Mandler et al. (1986), echoed by Squire et al. (1987), based on Shimamura's (1985, p. 507) claim that when "intertest biases" are present in an experiment, "the findings are equivocal." These researchers have argued that independence between recognition and primed fragment completion of the sort reported by Tulving et al. (1982) is not valid evidence because it may come about as a consequence of the priming that target words receive in the recognition test preceding the fragment-completion test.

It may be worth emphasizing that the finding of stochastic independence between recognition and primed fragment completion reported by Tulving et al. (1982) was originally interesting only because successive measures of episodic memory were widely and generally known to be dependent. Had it been known that successive measures of episodic memory were also independent, the results reported by Tulving et al. (1982) would not have been worth publishing. Thus, from the beginning, the important finding was differential dependency, albeit based on between-experiments comparisons. Shimamura's (1985) critique of the Tulving et al. (1982) conclusion ignored the implications of such a state of affairs.

The observation of different levels of dependence is not subject to possible problems of interpretation created by intertest biases, as was also acknowledged by Shimamura (1985). In the experiments using the triangulation method described here, we obtained differential dependency in four experimental conditions in which all variables that could have affected test priming, including the presentation of and judgments about test items in the first of two tests (recognition), were held constant. This means that both the original encoding of the studied material and whatever priming of it occurred during the first test were identical for the two conditions. It follows that the observed differences in dependence reflect the effect of something other than priming of the target words during the first test, namely different test instructions and, presumably, correspondingly different retrieval processes.

Recognition and Primed Word Completion Are Essentially Independent

We now turn to our finding that recognition and primed word-completion responses were essentially independent. In Experiments 1 and 2, the median Q was .093, and none of the four Q values was significantly different from 0. Nonetheless, there was a small positive dependence, and numerically three of the four obtained Q values were higher than 0. In other experiments in which stochastic independence between recognition and fragment completion has been demonstrated, the Q values have also been only near 0 and not 0. The two Q values in the Tulving et al. (1982, Table 1) experiment, which can be calculated from the old-word data, were .06 and .09. The Q values in the two corresponding conditions in an exact replication of the Tulving et al. experiment, briefly mentioned in Sloman, Hayman, Ohta, Law, and Tulving (1988), were .09 and .07. In the Light et al. (1986) experiment, the four Q values for recognition and study-and-test-primed fragment completion turn out to be .17, .37, .09, and .13. Such consistent findings in different experiments are a part of the empirical reality that theories of memory must address, even if the reality is represented only by repeated failures to reject the null hypothesis, mixed in with occasional observations of very low levels of dependency.

It probably would be unrealistic to expect perfect independence between tests on the same materials even if the tests were to tap quite different systems. Some observers may find it altogether surprising that subjects in experiments such as ours and those of Graf and Mandler (1984) and Graf et al. (1984) can adopt one or the other of two modes of retrieval, explicit or implicit. Although the results of these experiments clearly imply that subjects can do so, there is no compelling evidence that the system is free from "noise." It would be difficult to eliminate all the normal "cross-talk" between different systems that may result from subjects' use of deliberate strategies as well as from their ingrained habits of optimizing memory performance in all situations. Such cross-talk is reflected in subject and item correlations that are higher than 0. Only if a given task constituted a pure measure of a particular memory system—an assumption that seems unreasonable—would the expectation of perfect independence be realistic. This is why we prefer to think of the relation between recognition and primed fragment completion not as independent but as nearly or largely independent. It is possible, of course, that such near independence is a consequence of test priming, superimposed on two measures that are in fact correlated, as argued by Shimamura (1985) and others (e.g., Mandler et al., 1986). But until such time that the proponents of this hypothesis produce empirical support for it, it seems more prudent to think that the observation of independence tells us something other than the fact that test priming has occurred.

On the Nature of Priming

Given the predominantly methodological orientation of the research reported here, we need not dwell at length on theoretical implications of the findings. Different theoretical ideas concerning implicit memory have recently been reviewed by Richardson-Klavehn and Bjork (1988), Roediger, Weldon,

and Challis (in press), and Schacter (1987). All of the extant theories could explain the results of our experiments because these results represent replications and natural extensions of empirical findings already reported in the literature.

We prefer the view that the kind of learning that manifests itself in priming effects is mediated by a brain–cognition system different from that supporting episodic memory. We refer to the other hypothetical system as the QM (quasimemory) system. The functional properties of the QM system, and their differences from those of the episodic system, can be inferred from empirical findings of experiments comparing implicit and explicit memory tasks. Many such properties have been described by Roediger and his colleagues (in press) in the course of the development of arguments in support of the *transfer-appropriate procedural approach* that they favor. These properties include sensitivity to surface features of processed information and relative lack of sensitivity to semantic features. Other properties of the QM system have to do with the remarkable resistance of some QM information to forgetting (e.g., Sloman et al., 1988) and to interference (e.g., Graf & Schacter, 1987; Jacoby, 1983; Sloman et al., 1988). Less is known about the neural correlates of the QM system (but see Weiskrantz, 1987, for some speculations concerning neuroanatomical correlates of priming).

One of the major assumptions underlying the QM-system explanation of priming effects is that these effects are based on information and retrieval processes that differ from those that support phenomena of episodic memory. This assumption is supported by the fact that amnesic patients with brain damage that results in severely impaired episodic-memory capabilities frequently show normal or near-normal priming effects (e.g., Schacter, 1987; Schacter & Graf, 1986; Shimamura, 1986; Squire, Shimamura, & Graf, 1985; Warrington & Weiskrantz, 1970).

The notion of the QM system is compatible with the results of our experiments, as well as with those of most other experiments in which priming phenomena have been explored. It fits facts such as the sensitivity of priming effects to surface features of the stimulus materials, long-lasting priming effects, the early development of priming effects in children (e.g., Parkin & Streete, 1988), and resistance of priming effects to old age (e.g., Light et al., 1986). It is also compatible with the dissociation data from experiments with amnesic subjects mentioned earlier. The major shortcoming of the QM-system explanation of priming is its post hoc nature—it can be fitted to the data, but by itself it does not make any genuinely novel predictions.

Our concept of a separate QM system that normally interacts with other memory systems (Tulving, 1985a, 1987) in the processing of mnemonic information and the execution of various memory tasks is in good agreement with the transfer-appropriate procedural approach as advocated by Jacoby (1983), Roediger and Blaxton (1987a, 1987b), Roediger and Weldon (1987), Roediger, Weldon, and Stadler (1987), and elaborated by Roediger et al. (in press). The version of this approach described by Roediger et al. (in press) is based on four fundamental assumptions: (a) Outcomes of memory tests depend critically on the extent to which retrieval operations or information match those of encoding; (b) explicit (episodic) and implicit (nonepisodic) memory require

different retrieval operations; (c) retrieval of episodic information (referring to specific time and place in the rememberer's past) in tasks such as recall and recognition depends heavily on conceptually driven processes involving deeper, semantic encoding of the to-be-remembered material and the retrieval cues; and (d) retrieval of nonepisodic information (no such time-space reference) in tasks such as fragment and stem completion depends heavily on data-driven processes involving surface features of encoded material and the retrieval cues.

The transfer-appropriate procedural approach can account for many facts concerning priming, including those described in this article, but it has no natural explanations for findings such as the sharp dissociations between priming effects and episodic measures of memory observed in amnesia and normal development. The different patterns of performance obtained from normal subjects and amnesic patients cannot be attributed to differences in encoding and retrieval processes engendered by different task requirements or availability of one or another kind of information, because all these are held constant. Such different patterns of performance reflect the brain damage suffered by amnesics that differentially impairs the operation of brain–cognition systems whose overt manifestations are observed in explicit- and implicit-memory tasks.

Under the circumstances, an integration of processing and systems approaches to implicit memory seems to be called for. If the basic assumptions of the transfer-appropriate procedural approach are conceptualized in terms of functional properties of a distinct memory system, such as our hypothetical QM system, then the amnesia data would complement many findings with normal subjects. An approach that conjoins the main ideas of systems and processing views has distinct advantages over existing theoretical views. In advocating such an approach, we agree with similar sentiments expressed, although somewhat more tentatively, by Roediger et al. (in press) as well as with similar ideas mentioned by Richardson-Klavehn and Bjork (1988). The combined multiple-systems and processing framework would not only help us to understand phenomena of implicit and explicit memory but would also contribute to the developing enterprise of classification of learning and memory (Tulving, 1984, 1985b, 1986).

Conclusions

We conclude by summarizing the main points we have made.

1. The successive testing paradigm and contingency analyses of the data represent useful adjuncts to other methods of studying memory, especially if researchers' attention is diverted from the all-or-none contrast between dependence and independence to continuous measures reflecting different degrees of dependence.

2. The method of triangulation, in which the relation between two measures of memory is described in terms of their individual contingency relations to a third, common, reference measure, represents one way of at least partially solving the problem created by the existence of the effects of the first of two tests on the contingency relation between

them. The method of triangulation has three advantages over the practices of the past: (a) It conforms better to nature, inasmuch as it explicitly acknowledges that dependence is a variable quantity; (b) systematic variations in measured dependence as a function of experimental conditions constitute facts that are likely to be more tractable theoretically than simple claims of dependence or independence; and (c) systematic variations in dependence are less susceptible to, and less likely to evoke, claims regarding the artifactuality of the findings.

3. Yule's Q is a useful measure of dependency because of its natural range of -1 to $+1$ and because of ready availability of accepted tests of its significance as well as the significance of differences between Q values.

4. Our experiments have demonstrated that the contingency relation between retention performance on successive tests of studied material varies systematically with the nature of retrieval instructions given to subjects at test, with all other conditions of the experiment held constant. Fragment-cued recall showed greater dependence with recognition than did primed fragment completion, thus suggesting that the psychological and brain processes underlying the two kinds of retrieval are different. Such a contingent dissociation between cued recall and primed fragment completion corroborates corresponding findings of functional dissociations between measures of explicit (episodic) and implicit (QM) memory.

5. An adequate theoretical account of these findings is provided by an integration of the idea of multiple memory systems and the processing approach. Specifically, we propose that priming phenomena reflect the operation of a brain–cognition system—the *QM* system—that can operate, at least to some extent, independently of the episodic system and that, as other research has shown, can remain intact after brain damage causing amnesia. The functional properties of the QM system include sensitivity to surface features of processed information and relative lack of sensitivity to semantic features. Such an account also fits many other facts concerning priming and its relation to other forms of learning and memory.

References

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge, MA: MIT Press.

Broadbent, D. E., & Broadbent, M. H. (1975). The recognition of words which cannot be recalled. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (Vol. 5, pp. 575–590). New York: Academic Press.

Broadbent, D. E., & Broadbent, M. H. (1977). Effects of recognition on subsequent recall: Comments on "Determinants of recognition and recall: Accessibility and generation" by Rabinowitz, Mandler, and Patterson. *Journal of Experimental Psychology: General, 106,* 330–335.

Chandler, C. (1983). *Does retrieval strategy determine the relation between episodic recognition and semantic priming?* Unpublished master's thesis, University of Toronto, Toronto, Ontario, Canada.

Darley, C. F., & Murdock, B. B. (1971). Effects of prior free-recall testing on final recall and recognition. *Journal of Experimental Psychology, 91,* 66–73.

Eich, E. (1984). Memory for unattended events: Remembering with and without awareness. *Memory & Cognition, 12,* 105–111.

Fienberg, S. E. (1977). *The analysis of cross-classified categorical data.* Cambridge, MA: MIT Press.

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. (1988). Recognition failures and free-recall failures: Implications for the relation between recall and recognition. *Memory & Cognition, 16,* 446–451.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of American Statistical Association, 49,* 732–764.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior, 23,* 553–568.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 501–518.

Graf, P., & Schacter, D. L. (1987). Selective effects of interference on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 45–53.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 164–178.

Greeno, J. G., James, C. T., & DaPolito, F. J. (1971). A cognitive interpretation of negative transfer and forgetting of paired associates. *Journal of Verbal Learning and Verbal Behavior, 10,* 331–345.

Hilgard, E. R. (1955). Discussion of probabilistic functionalism. *Psychological Review, 62,* 226–228.

Hintzman, D. L. (1978). *The psychology of learning and memory.* San Francisco: Freeman.

Hintzman, D. L. (1980). Simpson's paradox and the analysis of memory retrieval. *Psychological Review, 87,* 398–410.

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

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 21–38.

Jacoby, L. L., & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology, 36,* 300–324.

Johnston, J. C. (1978). A test of the sophisticated guessing theory of word perception. *Cognitive Psychology, 10,* 123–153.

Light, L. L., Singh, A., & Capps, J. L. (1986). Dissociation of memory and awareness in young and older adults. *Journal of Clinical and Experimental Neuropsychology, 8,* 62–74.

Mandler, G., Graf, P., & Kraft, D. (1986) Activation and elaboration effects in recognition and word priming. *Quarterly Journal of Experimental Psychology, 38,* 645–662.

Mandler, G., Pearlstone, Z., & Koopmans, H. S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 8,* 410–423.

Martin, E. (1971). Verbal learning theory and independent retrieval phenomena. *Psychological Review, 78,* 314–332.

Martin, E. (1981). Simpson's paradox resolved: A reply to Hintzman. *Psychological Review, 88,* 372–374.

Mitchell, D. B., & Brown, A. S. (1988). Persistent repetition priming

in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 213–222.

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.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109–133.

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

Nilsson, L.-G., Law, J., & Tulving, E. (1988). Recognition failure of recallable unique names. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 266–277.

Ogilvie, J. C., Tulving, E., Paskowitz, S., & Jones, G. V. (1980). Three-dimensional memory traces: A model and its application to forgetting. *Journal of Verbal Learning and Verbal Behavior, 19,* 405–415.

Parkin, A., & Streete, S. (1988). Implicit and explicit memory in young children and adults. *British Journal of Psychology, 79,* 361–369.

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

Postman, L., Jenkins, W. O., & Postman, D. L. (1948). An experimental comparison of active recall and recognition. *American Journal of Psychology, 61,* 511–519.

Rabinowitz, J. C., Mandler, G., & Patterson, K. E. (1977). Determinants of recognition and recall: Accessibility and generation. *Journal of Experimental Psychology: General, 106,* 302–329.

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology, 39,* 475–543.

Roediger, H. L., III, & Blaxton, T. A. (1987a). Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. *Memory & Cognition, 15,* 379–388.

Roediger, H. L., III, & Blaxton, T. A. (1987b). Retrieval modes produce dissociations in memory for surface information. In D. Gorfein & R. R. Hoffman (Eds.), *Memory and cognitive processes: The Ebbinghaus centennial conference* (pp. 349–377). Hillsdale, NJ: Erlbaum.

Roediger, H. L., III, & Weldon, M. S. (1987). Reversing the picture superiority effect. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes: Theories, individual differences, and applications* (pp. 151–174). New York: Springer-Verlag.

Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (in press). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness.* Hillsdale, NJ: Erlbaum.

Roediger, H. L., III, Weldon, M. S., & Stadler, M. A. (1987, November). *Direct comparison of two implicit measures of retention.* Paper presented at the meeting of the Psychonomic Society, Seattle, WA.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 501–518.

Schacter, D. L., & Graf, P. (1986). Effects of elaborative processing on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 432–444.

Shimamura, A. P. (1985). Problems with the finding in stochastic independence as evidence for the independence of cognitive processes. *Bulletin of the Psychonomic Society, 23,* 506–508.

Shimamura, A. P. (1986). Priming effects in amnesia: Evidence for a dissociable memory function. *Quarterly Journal of Experimental Psychology, 38,* 619–644.

Sloman, S. A., Hayman, C. A. G., Ohta, N., Law, J., & Tulving, E.

(1988). Forgetting in primed fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 223–239.

Squire, L. R., Shimamura, A. P., & Graf, P. (1985). Independence of recognition memory and priming effects: A neuropsychological analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 37–44.

Squire, L. R., Shimamura, A. P., & Graf, P. (1987). Strength and duration of priming effects in normal subjects and amnesic patients. *Neuropsychologia, 25,* 195–210.

Tulving, E. (1983). *Elements of episodic memory.* Oxford: Clarendon Press.

Tulving, E. (1984). Multiple learning and memory systems. In K. M. J. Lagerspetz & P. Niemi (Eds.), *Psychology in the 1990's* (pp. 163–184). Amsterdam, The Netherlands: Elsevier.

Tulving, E. (1985a). How many memory systems are there? *American Psychologist, 40,* 385–398.

Tulving, E. (1985b). On the classification problem in learning and memory. In L-G. Nilsson & T. Archer (Eds.), *Perspectives in learning and memory* (pp. 67–94). Hillsdale, NJ: Erlbaum.

Tulving, E. (1986). What kind of a hypothesis is the distinction between episodic and semantic memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 307–311.

Tulving, E. (1987). Multiple memory systems and consciousness. *Human Neurobiology, 6,* 67–80.

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 336–342.

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

Tulving, E., & Watkins, M. J. (1975). Structure of memory traces. *Psychological Review, 82,* 261–275.

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

Upton, G. J. G. (1978). *The analysis of cross-tabulated data.* Chichester, NY: Wiley.

Wallace, W. P. (1978). Recognition failure of recallable words and recognizable words. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 441–452.

Warrington, E. K., & Weiskrantz, L. (1970). The amnesic syndrome: Consolidation or retrieval? *Nature, 228,* 628–630.

Watkins, M. J., & Todres, A. K. (1978). On the relation between recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 17,* 621–633.

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

Weiskrantz, L. (1987). Neuroanatomy of memory and amnesia: A case for multiple memory systems. *Human Neurobiology, 6,* 93–105.

Yule, G. U. (1900). On the association of attributes in statistics. *Philosophical Transactions of the Royal Society, London, Series A, 194,* 257–319.