EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY, 1993, 5 (4) 353–373

Stochastic Independence in the Recognition/ Identification Paradigm

Endel Tulving

Rotman Research Institute, Baycrest Centre, North York, and University of Toronto, Toronto, Ontario, Canada

C.A. Gordon Hayman

Department of Psychology, Lakehead University, Thunder Bay, Ontario, Canada

Experiments in which subjects study a set of items and then take two successive tests, a recognition test followed by a test of implicit retrieval of the same items (e.g. Tulving, Schacter, & Stark, 1982), have typically yielded data showing stochastic independence between the tests. Ostergaard (1992) has suggested that the procedure for assessing such independence should include an evaluation of the data against a model that assumes maximal dependence between the components of the test attributable to the study episode. In applying his Maximum Memory Dependence model to the published data, however, Ostergaard inexplicably used test-primed data for baseline measures, rather than non-primed data, and erroneously concluded that there was no clear evidence for independence. In this article, we show that when the standard baseline measures are used, the results of many experiments demonstrate stochastic independence.

INTRODUCTION

Contingency analyses of data from experiments using successive tests have provided useful information about the processes of interest to students of memory. The information is of a kind that is not available from other types

---

Requests for reprints should be addressed to Endel Tulving, Rotman Research Institute, Baycrest Centre, 3560 Bathurst Street, North York, Ontario M6A 2E1 *or* to Gordon Hayman, Department of Psychology, Lakehead University, Thunder Bay, Ontario P7B 5E1, Canada.

This research was supported by the Natural Sciences and Engineering Research Council of Canada Grants No. A8632 to Endel Tulving and No. OGP46356 to Gordon Hayman.

of experiments and other forms of analysis. The theoretically important finding from these experiments is that the degree of dependence between the tests varies systematically with the tests.

The data from successive-tests experiments are analysed in terms of the contingency relation between the outcomes of the two tests. Individual subject-items are entered into a four-fold contingency table in which successful and unsuccessful retrieval on one test are crossed with successful and unsuccessful retrieval on the other. The contingency relation (dependence, or association, or subject-item correlation) between the two tests can be expressed quantitatively in terms of a measure such as gamma or Yule's Q, which has a range of -1 to $+1$, and the reliability of differences between any given two Q-values can be assessed by means of the chi-square statistic (Hayman & Tulving, 1989a; Nelson, 1984).[1]

Available evidence suggests that the contingency relation between two tests depends systematically on (a) explicit versus implicit retrieval instructions, and (b) identical versus different cues on the two tests (Hayman & Tulving, 1989a). When both tests are explicit (Graf & Schacter, 1985; Schacter, 1990), the dependence between the tests may vary from very high to moderate, depending on the nature of the cues. It is very high, frequently approaching unity, with identical cues (e.g. Hayman & Tulving, 1989b; Wallace, 1978), as well as with cues that contain similar information in different formats, such as the cues in a recognition test and an extralist-cued recall test (e.g. LeVoi et al., 1983; Ogilvie, Tulving, Paskowitz, & Jones, 1980). The dependence is moderately positive, with Q-values near 0.5, with different cues (e.g. Hayman & Tulving, 1989a; 1989b; LeVoi et al., 1983; Ogilvie et al., 1980; Tulving & Watkins, 1975). It is also moderately positive in so-called recognition-failure experiments in which, too, the cues in successive tests are different (e.g. Bryant, 1991; Flexser & Tulving, 1978; Gardiner & Tulving, 1980; Neely & Payne, 1983; Nilsson, Law, & Tulving, 1988; Tulving & Wiseman, 1975; Watkins & Tulving, 1975).

When one of the two successive tests is an implicit one, or when both are, the contingency relation may vary from high positive to virtual independence. Identical cues in the two tests produce a high degree of dependence (Hayman & Tulving, 1989b; Witherspoon & Moscovitch, 1989), whereas different cues usually produce very low degrees of dependence that frequently are indistinguishable from stochastic independence

---

[1]Hayman and Tulving's (1989a, p. 231) equation 5, which describes the formula for testing the significance of difference between two 2 × 2 tables, is incorrect. The log transformations of odds ratios in equation 5 should be ln (natural logarithm) transformations. The correct equation 5 is: $\chi^2 = (\ln C_1 - \ln C_2)^2/[V(C_1 + V(C_2)]$.

(e.g. Hayman & Tulving, 1989a; 1989b; Tulving, Schacter, & Stark, 1982; Tulving, Hayman, & Macdonald, 1991; Witherspoon & Moscovitch, 1989). Because this differential dependence (high or low degrees of dependence, depending upon tests) occurs when subjects and nominal target items are held constant, the findings cannot be attributed to subject and item correlations (cf. Flexser, 1981; 1991; Hintzman & Hartry, 1990).

RECOGNITION/IDENTIFICATION PARADIGM

The particular version of the successive-tests paradigm on which we focus in this article is one in which subjects, after studying a list of items, are given an explicit recognition test for these items, followed by an implicit retrieval test requiring them to identify same items on the basis of impoverished perceptual descriptions of the items. We will refer to this class of experiments as "recognition/identification" experiments, to distinguish them from other kinds of experiments involving successive tests, such as recognition-failure experiments (e.g. Nilsson & Gardiner, 1993; Watkins & Tulving, 1975).

An early recognition/identification experiment was reported by Tulving et al. (1982). Subjects were exposed to a list of 96 words on a single study trial, and then, either 1 h later or 7 days later, took two tests.[2] The first was the standard yes/no recognition test in which the subjects' task was to determine whether or not they had seen a given test word in the study phase of the experiment. In the second test, subjects were presented with graphemic fragments of words, such as A--AS--N and U--VE--E, and asked to identify the words by completing the fragments. They were told that some of the words had appeared in the earlier study list whereas others had not, and that their task was to complete each fragment with a meaningful word regardless of the word's earlier appearance in the experiment. Thus, the instructions given to the subjects at test were for what later came to be referred to as "implicit retrieval" (Graf & Schacter, 1985). In Tulving and co-workers' (1982) experiment, each fragment allowed only a single solution, although the subjects were not apprised of that fact.

The fragment-completion results of Tulving and co-workers' (1982) experiment showed priming: probability of fragment-completion was higher for study-list words that a subject had encountered in the experiment than for words not encountered in the experiment before the fragment-completion test. An important feature of this priming, with

[2]The design of the experiment was somewhat more complex than indicated here. In this article we are concerned with essentially one half of the design: recognition test preceding the fragment-completion test.

respect to the issue under scrutiny in this article, was that the exact circumstances of the encounter with the target words did not seem to matter. Priming occurred for words that the subjects had intentionally studied in the original list, as well as for words that had served as distractor items in the earlier recognition test. Specifically, the mean fragment-completion probabilities were 0.31 for the non-primed ("new") words (words neither in the study list nor in the recognition test), 0.46 for the "study-primed" words (words in the study list but not in the recognition test), 0.54 for "test-primed" words (words not in the study list but appearing as distractor items in the recognition test) and 0.65 for "study-and-test primed" words (words in the study list that also appeared as "old" items in the recognition test). These data were highlighted in Tulving and co-workers' (1982) article through their graphic presentation in figure 2, in which probability of fragment completion was shown for "unprimed" and "three kinds of primed words". The magnitude of the priming effect was larger for the recognition-test distractors (test-primed items) than study-list words, presumably because the retention interval was shorter for the former than the latter.[3]

A more interesting observation than that of the mere priming effect in Tulving and co-workers' (1982) experiment was the finding of stochastic independence between recognition and fragment completion: The probability of successful fragment completion in the second test was statistically non-discriminable for the words that the subjects had positively identified as study-list items and for the words that the subjects claimed they did not recognise as "old". Stochastic independence held for both study-list words and recognition-test distractors. That is, subsequent fragment completion was essentially the same for the items representing "hits" and those representing "misses", and also the same, albeit at a different level, for items representing "false alarms" and those representing "correct rejections".

At about the same time that Tulving and co-workers' (1982) results were published, a similar observation of stochastic independence between recognition and primed perceptual identification of words was reported by Jacoby and Witherspoon (1982). Together, the two sets of results were interesting, because they seemed to have produced a glaring exception to a highly consistent pattern of data known at the time. Before 1982, contingency analyses of performance on successive tests of the same target

[3]The results of many subsequent experiments have converged on the empirical generalisation that perceptual priming of the kind demonstrated in fragment completion depends primarily on the perceptual encounter with the target word rather than on the characteristics of the encounter. For reviews, see Challis and Brodbeck (1992), Roediger and Blaxton (1987), Roediger and McDermott (1993), Schacter (1990) and Tulving and Schacter (1990).

items had invariably demonstrated positive dependence or association between the tests (e.g. Brown, 1923; Flexser & Tulving, 1978; Gardiner & Tulving, 1980; Ogilvie et al., 1980; Postman, Jenkins, & Postman, 1948; Tulving & Watkins, 1975; Tulving & Wiseman, 1975; Wallace, 1978; Watkins & Tulving, 1975). Against the backdrop of such consistent evidence, the observed stochastic independence was surprising, and therefore worthy of theoretical attention. Here, for the first time, were data showing that two "memory effects" (recognition and priming), originating in one and the same encoding episode, behaved rather differently from all the previously reported comparisons—they were dissociated rather than associated.

A number of subsequent experiments have reported findings of stochastic independence (e.g. Hayman & Tulving, 1989a; Musen & Treisman, 1990; Schacter et al., 1991; Snodgrass & Feenan, 1990), although failures to replicate have also been reported (Hintzman & Hartry, 1990). At the empirical level, as already mentioned, most of the relevant data can be conveniently summarised in terms of two independent variables: (a) type of test instructions (explicit *vs* implicit) and (b) similarity relations between cues in the two tests (same *vs* different). At the theoretical level, the dominant themes in the interpretations of the findings have been those of multiple memory systems (e.g. Schacter, 1990; Tulving, 1983; Tulving & Schacter, 1990; Tulving et al., 1982) and process dissociations (e.g. Graf & Ryan, 1990; Jacoby, 1983; Roediger, Weldon, & Challis, 1989).

Recently, Ostergaard (1992) has expressed scepticism about the validity of the reports of stochastic independence in the recognition/identification paradigm and, in a sweeping generalisation, has called into question the general usefulness of contingency analyses in memory research. This article is concerned with Ostergaard's ideas and conclusions. We make three main points. First, Ostergaard's (1992) theoretical criticism of the procedures that have been used in the past to evaluate the hypothesis of stochastic independence is well taken, valid and useful. His Maximum Memory Dependence (MMD) model constitutes a definite methodological improvement. Second, for reasons unknown and not discussed, he adopted incorrect baseline measures in his calculations. As a consequence, he arrived at an erroneous overall conclusion. Third, when we apply Ostergaard's (1992) model to the extant data correctly, the results show that a large majority of the results of recognition/identification experiments conform much more closely to stochastic independence than dependence.

OSTERGAARD'S MMD MODEL

Ostergaard (1992) drew attention to a perceived shortcoming in the evaluation of the data that have been claimed to demonstrate stochastic independence. His argument can be summarised as follows. The conventional practice has been to compare the observed dependency between the two tests with the hypothesised state of stochastic independence, and then, when the observed dependency is small or not reliably different from the null hypothesis, to accept the null hypothesis of independence. This procedure is inadequate, however, because it is logically possible for the data from a given experiment to be such that the null hypothesis of independence cannot be rejected even if in fact there is a great deal of dependence present. This kind of outcome may occur in situations in which the "memory" component of the measured performance—the component attributable to the experimentally controlled encounter with the target items—is small, and in which, therefore, the post-encounter assessment of the degree of association between the two tests is largely determined by the pre-encounter association between the tests. Ostergaard (1992) presented actual numerical examples of these kinds of outcomes.

Because of the existence of these kinds of possibilities—maximum dependence that is not statistically different from independence because of small "memory" components of measured performance—Ostergaard suggested that the data from any given experiment should be compared with two models, one of independence and the other of maximum possible dependence. The conclusion drawn about the data should then be based on the outcomes of both comparisons.

Ostergaard proposed a Maximum Memory Dependency model that allows one to estimate the degree of maximum dependence that is possible between the outcomes of the two tests in any given situation. According to the model, the observed performance on each test, as well as the contingency relation between them, can be partitioned into two components: (a) pre-experimentally determined (the baseline) and (b) experimentally determined (the "memory" component). Each of the two tests comprising a recognition/identification experiment has its own "memory" component. In the recognition test, it is the "corrected recognition" score (the hit rate minus the false alarm rate). In the identification test, the "memory" component is the "priming effect" (the difference between identification probabilities for primed and non-primed items). The theoretically interesting question concerns the relation between these two "memory" components of the measures. The pre-experimentally determined relation between the two baseline components, estimated by false alarms in recognition and completion of fragments of non-primed words in the completion test, is not at issue.

The baseline performance is determined by variables that are usually held constant in the experiment. The natural expectation is that the baseline performances of the two tests are stochastically independent. The dependence between the "memory" components of the two test performances is merely superimposed on the relation of independence between the two baseline performances. The actually observed dependence between the two measures of performance varies both with the baseline performance and the "memory" components of the two measures. So does the maximum possible value of the dependence. It, too, is a function of the baseline performance of the two tests and the magnitude of the "memory" components. The important point to note is that one cannot assume that this maximum dependence is equal to unity. Frequently it may be less, and sometimes it is possible that the maximum possible dependence is indistinguishable from stochastic independence.

Ostergaard (1992) described his MMD model with respect to the general situation involving any two tests, T1 and T2. We adopt his description here for the typical recognition/identification experiment, in which the first test is recognition (Rn) and the second test that of fragment completion (FC).

Consider a situation in which the "memory" effect is smaller in fragment completion than it is in recognition:

$$p(M_{fc}) = p(FC_o) - p(FC_n) < p(M_{rn}) = p(H) - p(FA) \qquad (1)$$

where FC_o = fragment completion of "old" items, encountered in the experiment; FC_n = fragment completion of items not encountered earlier in the experiment; H = hit rate; FA = false alarm rate; M_{rn} = "memory" component of recognition; and M_{fc} = "memory" component of fragment completion.

Then, according to the MMD model (Ostergaard, 1992, p. 415), the maximum possible value of the joint probability of recognition and fragment completion of studied target items is the sum of the "memory" component in fragment completion and the product of the baseline in fragment completion and the hit rate in recognition:

$$\max p(Rn, FC) = p(FC_o) - p(FC_n) + p(FC_n)p(H) \qquad (2)$$

Note that in the extreme case in which the "memory" component of FC is zero (i.e. there is no priming), the maximum possible value of the joint probability of recognition and fragment completion of studied target items is equal to the product of fragment completion of non-primed items and the recognition hit rate. Such a state of affairs implies that the smaller the priming effect, the more likely it is that the relation between recognition and fragment completion of experimentally presented items is indistinguishable from the pre-experimental relation between the tests, that is, that it appears to be one of stochastic independence. It is this kind of an artifact that Ostergaard (1992) was concerned with in his article.

Using the MMD model, Ostergaard (1992) calculated the estimated levels of maximum memory dependence for 15 experimental conditions in five published experiments, and then compared the experimentally observed contingency (joint probabilities and Q-values) against those predicted by (a) the model of stochastic independence and (b) the MMD model. He summarised the results of his calculations in his table 4 (Ostergaard, 1992, p. 418). His putative critical finding was that, in all 15 cases, the joint probabilities (of success on both tests) predicted by both models fell within the 95% confidence interval of the observed joint probabilities. Moreover, in 8 out of 15 cases, Yule's Q of the 2×2 contingency table corresponding to the MMD model did not differ reliably from stochastic independence.

Ostergaard concluded that the reported data were inconclusive, that the data that had been reported in support of the hypothesis of stochastic independence were also compatible with the hypothesis of maximum "memory" dependence, and that under the circumstances the earlier claims that the data reflected stochastic independence had to be rejected. The major problem seemed to lie in the small "memory" components of the tests, usually reflecting small amounts of priming that had been observed in the experiments.

Taken at their face value, these conclusions are clearly highly damaging to the earlier declarations that the tests in recognition/identification experiments are stochastically independent (e.g. Hayman & Tulving, 1989a; Tulving et al., 1982). But, as sometimes happens, in this case too the appearance turns out to be deceiving, because of Ostergaard's (1992) choice of incorrect data.

OSTERGAARD'S ERROR

Ever since experiments on "repetition" priming have been done in cognitive psychology, priming effects have always been expressed in terms of the difference between primed ("old" or "studied") and non-primed ("new" or "non-studied") items. Primed items are those that have been previously encountered in the experiment, whereas non-primed items are those that appear in the experiment only at test. Such a definition of priming effects, in terms of "primed performance" and the "non-primed baseline", has been universally accepted (e.g. Hintzman & Hartry, 1990; Richardson-Klavehn & Bjork, 1988; Snodgrass & Feenan, 1990). No exceptions to the practice have ever been reported or proposed. Ostergaard (1992), however, did just that. Without providing any reason for his choice, he adopted a different baseline measure for his calculations. We think it was an error.

The baseline performance in the MMD model is defined in terms of the conventional baseline performance, that is, in terms of performance on items not previously encountered in the experiment. In most of his actual calculations, however, Ostergaard (1992) selected as his baseline the performance on items that in fact had been previously encountered in the experiment. He confused "new" items with "non-primed" items in most of the experiments he scrutinised.

In the recognition/identification paradigm, in which the two tests are given successively, previously non-encountered items differ for the two tests. In the recognition test, the non-encountered items are those that did not occur in the study list. That is, in the recognition test, non-studied items are "new" items, and vice versa. In the (following) identification test, however, non-studied items are not necessarily "new". Indeed, frequently in previous experiments, items not in the study list have been "old" by the time they appear in the identification test, by virtue of their having appeared as distractor items in the (earlier) recognition test. In this kind of situation, non-studied items appearing in the identification test are said to have been "test-primed". The important point is that with respect to the test on which priming is measured, these items are neither "new" nor "non-primed". It is these items, however, that Ostergaard chose for his baseline data in most of the experiments he examined.

The fact that priming occurs as readily for the items encountered only in the recognition test as it does for items studied in the list, first reported by Tulving et al. (1982), is by now well established. As already mentioned, in Tulving and co-workers' (1982) original experiment, items appearing only in the study list and those appearing only in the recognition test showed priming. Indeed, the completion rate was numerically higher for the "test-primed" items (0.54) than for the "study-primed" items (0.46). Both were considerably higher than the completion rate for the truly "new" (non-primed) items, which in Tulving and co-workers' (1982) experiment was 0.31.

Tulving and co-workers' (1982) experiment was one of those included in Ostergaard's database. For the purposes of the calculations of the MMD model data for the experiment, Ostergaard should have selected the non-primed performance (0.31) as the baseline, but he selected the test-primed performance (0.54) instead. The crucial consequence of this choice was that he estimated the "memory" component in this case as 0.11 (0.65 − 0.54), when in fact the true "memory" component was three times as large at 0.34 (0.65 − 0.31).

A CONCRETE EXAMPLE

To facilitate the comprehensibility of our argument, we next work through a concrete numerical example for one of the cases in Ostergaard's (1992) database for which the maximum memory dependence calculations are shown in his table 4. The example involves data from the HiC ("high constraint") condition in experiment 2, summarised in his tables 3 and 4, reported by Hayman and Tulving (1989a).

As shown in Hayman and Tulving's (1989a) table 4, the hit rate, p(H), in this condition was 0.645. As reported in Hayman and Tulving's (1989a) table 3, the probability of fragment completion of studied words, $p(FC_o)$, was 0.493. Also in table 3, two probabilities were given for the fragment completion of non-studied words, $p(FC_n)$. One (0.410) was for the "new" words tested in recognition (i.e. test-primed words) and the other (0.121) was for the "new" words not tested in recognition (i.e. non-primed words). The proportion of 0.121 is the true estimate of the baseline, whereas the proportion of 0.410 represents the combined effects of the baseline and the priming effect attributable to the appearance of the words in the recognition test.

Using the test-primed data (0.410) as the baseline, Ostergaard (1992, table 4) calculated the priming effect as $0.493 - 0.410 = 0.083$, and, applying the formula in equation 2, reached the conclusion that max p(Rn, FC) in this situation was 0.346:

$$\max p(Rn, FC) = 0.493 - 0.410 + (0.410)(0.645) = 0.346 \qquad (3)$$

This joint probability, 0.346, predicted on the basis of the MMD model, is indistinguishable from the observed data, $p(Rn, FC) = 0.344$. It is redundant to note that Yule's Q expressing the degree of dependence between recognition and primed fragment completion for the MMD model data is 0.251, which is practically identical with the Q of actual data, 0.230 ($\chi^2 = 0.02$). Because the contingency data showed the maximum dependence possible under the circumstances, the conclusion appears to be inescapable that in this particular case the two "memory" components—the priming effect and the corrected recognition—may be perfectly correlated.

The conclusion, however, is false, because of the inappropriate baseline data that Ostergaard used. He should have selected the non-primed "new" data (0.121) for his calculations shown in equation 3, instead of selecting the test-primed data (0.410). Had he done so, he would have obtained the value of 0.449 as the estimate of the maximum memory dependence in this condition:

$$\max p(Rn, FC) = 0.493 - 0.121 + (0.121)(0.645) = 0.449 \qquad (4)$$

This correct value of 0.449 is considerably higher than the observed p(Rn, FC) of 0.344. The corresponding values of Yule's Q are also very different: $Q = 0.89$ for the MMD model and $Q = 0.23$ for the experimental data. The chi-square expressing the significance of the difference between these two Q-values is 27.7 ($P < 0.001$). At the same time, the observed p(Rn, FC) of 0.344 is not significantly different from the value predicted by the model of stochastic independence, namely 0.317 ($\chi^2 = 3.13$, $P > 0.05$). Thus, while the hypothesis of maximum "memory" dependence can be safely rejected, the hypothesis of stochastic independence cannot.

Ostergaard (1992) made similar mistakes in most of the 15 sets of data that he summarised in his table 4. Although he did not provide explicit numerical information about the new-item baselines in his article, the calculations reported in his table 4 allow the inference that just about in all cases where it was possible to do so, he used test-primed rather than non-primed data as the baseline.[4] Ostergaard's (1992) untoward conclusions regarding stochastic independence are based on his choice of the wrong baseline data. When the correct baseline is used, the total picture changes drastically from that depicted by him.

In Table 1 we present the correct calculations for the data from the 15 experimental conditions that Ostergaard (1992) summarised in his table 4. Table 1 shows joint probabilities, observed dependence and predicted maximum dependence using Ostergaard's formula for estimating maximum dependence, but substituting untested new items (non-primed items) for test-primed new items. The data are presented in the same format used by Ostergaard, with two additional sets of statistics for each experimental condition: We have added the estimated 95% confidence interval for the observed Yule's Q as well as the marginal probabilities of "old" recognition, "old" completion and "new" completion responses.

The data in Table 1 make two points. First, stochastic independence was expected in all 15 conditions summarised in Table 1, and in keeping with this expectation, the observed relation was not significantly different from that predicted by stochastic independence in 14 out of 15 cases. The sole exception is one of the four conditions in Light, Singh and Capps (1986), in which the observed Yule's Q (0.370) falls outside the 95% confidence interval. Second, the observed dependence in Table 1 is significantly lower than the estimate of maximum dependence in 13 of the 15 conditions. The two exceptions are experiments 2 and 3 (with "elaborate" instructions) from Schacter, Cooper and Delaney (1990), where completion rates for

[4]It is worth pointing out that in the single concrete example that Ostergaard presented in his article to illustrate how the MMD calculations work (Ostergaard, 1992, pp. 415–417), he did use the conventional, correct, non-primed baseline. This fact suggests that (a) he chose test-primed baselines for other experiments unwittingly, and (b) that the original reviewers of the article in the *Journal of Experimental Psychology: Learning, Memory and Cognition* were misled by the use of the correct baseline in the example into believing that Ostergaard's other calculations were also correct.

TABLE 1
Observed Joint Probabilities Compared with Joint Probabilities Expected from Stochastic Independence and from the Model of Maximum Dependence

	Joint Probability				*Yule's Q*				*Marginal ppn*		
	OBS	*95% CL*	*SI*	*M × D*	*Q/obs*	*CI*	*Q/M × D*	χ^2	*RN*	*Cm*	*CmN*
Tulving, Schacter and Stark (1982)											
RN–FC 1 h	0.51	0.47–0.55	0.51	0.58	0.059	0.20	0.738	42.21	0.78	0.65	0.30
RN–FC 7 days	0.36	0.35–0.43	0.38	0.52	0.090	0.17	0.900	85.73	0.58	0.65	0.31
Light, Singh and Capps (1986)											
RN–FC Y Imm	0.64	0.59–0.70	0.63	0.68	0.165	0.37	0.752	9.79	0.90	0.71	0.30
RN–FC Y 7 days	0.51	0.45–0.56	0.47	0.59	0.370	0.21	0.872	21.14	0.68	0.69	0.30
RN–FC O Imm	0.57	0.51–0.62	0.56	0.60	0.092	0.34	0.666	11.20	0.88	0.64	0.30
RN–FC O 7 days	0.40	0.35–0.46	0.39	0.48	0.133	0.23	0.712	24.62	0.65	0.60	0.33
Hayman and Tulving (1989a)											
Exp. 1 Cmpl HiC	0.46	0.40–0.52	0.44	0.49	0.178	0.29	0.527	5.58	0.79	0.56	0.34
Exp. 1 Cmpl LoC	0.29	0.23–0.35	0.30	0.36	−0.160	0.27	0.570	27.94	0.73	0.42	0.21
Exp. 2 Cmpl HiC	0.34	0.29–0.40	0.32	0.45	0.230	0.24	0.887	27.66	0.65	0.49	0.12
Exp. 2 Cmpl LoC	0.23	0.18–0.28	0.23	0.30	0.001	0.26	0.636	22.96	0.61	0.38	0.18

Schacter, Cooper and Delaney (1990)											
Exp. 1	0.55	0.48–0.62	0.54	0.60	0.105	0.38	0.677	9.46	0.67	0.81	0.63
Exp. 2	0.55	0.48–0.62	0.52	0.55	0.303	0.32	0.293	0.00	0.69	0.76	0.68
Exp. 3 Hght/Wdth	0.54	0.47–0.61	0.52	0.58	0.286	0.35	0.714	6.52	0.64	0.81	0.63
Exp. 3 Elaborate	0.66	0.60–0.73	0.65	0.66	0.236	0.41	0.285	0.06	0.83	0.78	0.69
Schacter et al. (1991)											
Exp. 1	0.48	0.41–0.54	0.47	0.52	0.075	0.32	0.500	6.82	0.65	0.72	0.57

Note: OBS = observed joint probability; 95% CL = confidence limits of OBS; SI = stochastic independence; M × D = maximum dependence; Q/obs = Yule's Q for observed contingency table; CI = 95% confidence interval for observed Yule's Q; Q/M × D = Yule's Q for estimated maximum-dependence contingency table; χ^2 = chi-square test of the difference between contingency for observed and estimated-maximum dependence; RN = recognition of studied items; Cm = completion for studied items; CmN = completion for new (non-primed) items; RN–FC = RN followed by fragment completion; Y = young; O = old; Imm = immediate test; Cmpl = completion; HiC = high constraint on fragment completion; LoC = low constraint on fragment completion. Yule's Q-values were calculated from contingency tables based on the data reported by the authors and on tables predicted by the hypothesis of maximum dependence. Chi-square values in excess of 3.84 indicate that the corresponding observed Yule's Q-value differs from that estimated for maximum dependence. All significance tests were based on procedures recommended by Hayman and Tulving (1989a). Hayman and Tulving (1989a, exp. 1) did not test fragments for non-primed items. An estimate for the new items was derived from the probability of a non-target word response in the study-only condition.

unprimed items were very similar to those for primed items, that is, where priming effects were very small. Thus, the bulk of the data in Table 1 is clearly at variance with Ostergaard's model's predictions. In a large majority of published experiments, recognition and primed identification of target items are stochastically independent. Exceptions—cases where the outcome is inconclusive—seem to be largely limited to situations in which the observed priming was relatively small.

AN ADDITIONAL TEST OF THE MMD MODEL

We emphasise that we are in no way criticising Ostergaard's (1992) MMD model. In our opinion, it represents a useful methodological advance for the analysis of contingency data. Our criticism is directed solely at Ostergaard's (1992) use of inappropriate data.

We next present some data that further illustrate the usefulness of Ostergaard's model. These data come from an extensive case experiment involving the amnesic subject K.C. (Tulving et al., 1991). In the experiment, K.C. was repeatedly exposed to a set of 64 critical target words, and repeatedly tested over a period of many weeks for his ability to produce these target words to either conceptual cues (pictures or sentence frames, or both) or perceptual cues (word fragments). Thus, for example, K.C. learned the sentence, LIFEGUARD BOUGHT BATHROBE, among other comparable sentences in the set of 64, and then, in test sessions, he was asked to produce the target word BATHROBE to cues such as sentence frames (LIFEGUARD BOUGHT ???) or fragments (B---RO-E). In a given test session, only one type of cue would be used. The target words in all test sessions were always identical, consisting of the 64 words in the critical set. Across different sessions, separated by days or weeks, the relation of K.C.'s performance on one test could be compared with his performance on another test, and the contingency relation between the tests measured. K.C.'s production and completion rates were also available for non-studied words.

The data provided by K.C. are different from the data that normal subjects would produce in comparable experiments, for three reasons. First, because K.C. has no functional episodic memory, he cannot rely on his episodic recollection of what he has learned. The conceptually cued (sentence frames) and perceptually cued (word fragments) tests tap "pure" implicit retrieval, "uncontaminated" by any "explicit strategies" that may affect the performance in comparable tests given to normal subjects. Second, the long inter-test intervals, ranging up to 91 days, would be expected to minimise the "contaminating" effects of one test on the other, a concern about the successive-tests method expressed by many (e.g.

Mandler, Graf, & Kraft, 1986; Richardson-Klavehn & Bjork, 1988; Shimamura, 1985). Third, successive tests in the experiment measured the "steady-state" of the information acquired by K.C., inasmuch as there was little change in his overall performance over many testing sessions.

The data summarised in Table 2 in this article are imported from table 6 of Tulving et al. (1991). They express the extent to which test performance in one test is dependent on test performance in another test in which the same cues are used. Over a wide range of intervals between the tests (from 21 to 80 days, as described in Tulving and co-workers' article), and regardless of the type of cues used (word fragments, sentence frames or pictures), these data revealed uniformly high levels of Q-values.

The important point made by the data in Table 2 is that, by and large, they conform very closely to the predictions made by the MMD model. Thus, the mean of the 10 observed values of joint probabilities of successful performance on Test 1 and Test 2 was 0.53 for the actual data and 0.54 for the MMD model-predicted data. Similarly, the 10 Q-values of the contingency tables yielded a mean of 0.89 for the actual data and 0.88 for the MMD model. Finally, the model-predicted joint probabilities of successful Test 1 and Test 2 performances were comfortably within the 95% confidence intervals of the observed probabilities in all 10 cases. One could hardly ask for better agreement between actual data and those predicted by a model.

Ostergaard's (1992) model helps us to interpret the data summarised in Table 2: The "memory" components of successive tests, in which the same cues are directed at the same target items, under the same retrieval instructions, are essentially perfectly correlated. The pattern of data depicted by the 10 conditions summarised in Table 2 shows that the dependence between the tests was the maximum possible under the circumstances.

The significance of the data in Table 2, and their interpretation in light of Ostergaard's model, is underscored by a different set of data, shown in Table 3. These data have been taken from table 8 of Tulving et al. (1991). They comprise contingency analyses of nine pairs of successive tests in which different cues were used between the tests. Thus, in each of the nine comparisons, fragment cues were used in one test session for the 64 targets, and sentence frames were used in the other test session. The target words required as responses were again identical in the two tests.

Table 3 shows that when different cues are used in successive implicit tests of retrieval, the performance on one test is stochastically independent of the other test. In all nine comparisons, the relation between perceptual priming and semantic learning is close to that predicted by stochastic independence, and highly significantly different from that predicted by the maximum dependence hypothesis. The mean Q-value of the experimental

TABLE 2
Estimates of Maximum Dependence for Within-Task Performance from Table 6 of Tulving et al. (1991)

		OBS	*95% CL*	*SI*	*M × D*	*Q/obs*	*CI*	*Q/M × D*	χ^2	*TO1*	*TN1*	*TO2*	*TN2*
OF	1	0.56	0.44–0.68	0.42	0.57	0.91	0.11	0.94	0.15	0.63	0.16	0.67	0.28
	2	0.59	0.47–0.71	0.44	0.58	0.96	0.07	0.92	1.15	0.63	0.16	0.70	0.28
	3	0.63	0.51–0.74	0.47	0.59	0.95	0.07	0.86	7.15	0.67	0.28	0.70	0.28
NF		0.28	0.17–0.39	0.17	0.23	0.76	0.24	0.44	6.64	0.42	0.24	0.41	0.31
PSe		0.39	0.27–0.51	0.26	0.48	0.81	0.20	1.00	3.46	0.48	0.00	0.55	0.00
Se	1	0.36	0.24–0.48	0.22	0.44	0.86	0.16	0.99	2.85	0.45	0.03	0.48	0.03
	2	0.41	0.29–0.53	0.27	0.44	0.89	0.15	0.98	1.50	0.45	0.03	0.59	0.09
	3	0.42	0.30–0.54	0.29	0.47	0.86	0.16	0.98	2.12	0.48	0.03	0.59	0.09
PSeOF		0.84	0.76–0.93	0.78	0.82	0.97	0.06	0.88	7.05	0.86	0.38	0.91	0.47
PSeNF		0.83	0.74–0.92	0.75	0.79	0.96	0.07	0.76	32.20	0.88	0.56	0.86	0.56
Mean						0.89		0.88					

Note: OBS = observed joint probability; 95% CL = confidence limits of OBS; SI = stochastic independence; M × D = maximum dependence; Q/obs = Yule's Q for observed contingency table; CI = 95% confidence interval for observed Yule's Q; Q/M × D = Yule's Q for estimated maximum-dependence contingency table; χ^2 = chi-square test of the difference between contingency for observed and estimated-maximum dependence; TO1 = old on test 1; TN1 = new on test 1; TO2 = old on test 2; TN2 = new on test 2; OF = old fragment; NF = new fragment; P = picture; Se = sentence frame. Chi-square values in excess of 3.84 indicate that the corresponding observed Yule's Q-value differs from that estimated for maximum dependence.

TABLE 3

Estimates of Maximum Dependence for Old Fragments Versus Sentence Frame from Table 8 of Tulving et al. (1991)

Session	*OBS*	*95% CL*	*SI*	*M × D*	*Q/obs*	*CI*	*Q/M × D*	χ^2	*TO1*	*TN1*	*TO2*	*TN2*
7-10	0.30	0.19–0.41	0.28	0.44	0.12	0.50	0.98	11.08	0.63	0.16	0.45	0.03
7-13	0.30	0.19–0.41	0.30	0.47	−0.05	0.51	0.98	15.88	0.63	0.16	0.48	0.03
7-22	0.36	0.24–0.48	0.37	0.56	−0.11	0.51	0.98	17.23	0.63	0.16	0.59	0.09
15-10	0.28	0.17–0.39	0.30	0.44	−0.21	0.50	0.97	21.26	0.67	0.28	0.45	0.03
15-13	0.33	0.21–0.44	0.33	0.47	0.02	0.52	0.98	12.71	0.67	0.28	0.48	0.03
15-22	0.41	0.29–0.53	0.40	0.56	0.06	0.53	0.96	11.17	0.67	0.28	0.59	0.09
19-10	0.33	0.21–0.44	0.32	0.44	0.09	0.54	0.97	10.14	0.70	0.56	0.45	0.03
19-13	0.33	0.21–0.44	0.34	0.48	−0.12	0.53	0.97	16.20	0.70	0.56	0.48	0.03
19-22	0.41	0.29–0.53	0.42	0.57	−0.12	0.54	0.95	14.80	0.70	0.56	0.59	0.09
Mean					−0.04		0.97					

Note: OBS = observed joint probability; 95% CL = confidence limits of OBS; SI = stochastic independence; M × D = maximum dependence; Q/obs = Yule's Q for observed contingency table; CI = 95% confidence interval for observed Yule's Q; Q/M × D = Yule's Q for estimated maximum-dependence contingency table; χ^2 = chi-square test of the difference between contingency for observed and estimated-maximum dependence; TO1 = old on test 1; TN1 = new on test 1; TO2 = old on test 2; TN2 = new on test 2. Chi-square values in excess of 3.84 indicate that the corresponding observed Yule's Q-value differs from that estimated for maximum dependence.

data (Q = −0.04) is drastically different from the mean Q-value predicted by the MMD model (Q = 0.97).

We have pointed out earlier (Tulving et al., 1991) that the published data suggest that in situations in which repeated tests, directed at the same nominal targets, involve different cues, and at least one of the tests is one of implicit retrieval, the outcomes of the tests yield stochastic independence. The data summarised in Table 3 forcefully corroborate this generalisation. Ostergaard's (1992) model has helped us to rule out at least one alternative interpretation of the data, namely that the failure to observe dependence between perceptual priming and semantic learning is attributable to the insufficient magnitude of the "memory" components of the measured performance.

SUMMARY

No existing method for the study of memory is perfect, and all can benefit from improvement. Whenever such improvement occurs, we can call it progress. Ostergaard (1992) used his MMD model for the purpose of re-examining some 15 sets of data from a variety of recognition/identification experiments, and on the basis of his findings drew a broad, damning conclusion:

> Taken together with other problems associated with contingency analysis such as Simpson's paradox (Hintzman, 1980), intertest biases (Shimamura, 1985), and item and subjects characteristics and their interaction (Hintzman & Hartry, 1990), it is becoming increasingly clear that, unless the technique can be greatly improved, contingency analyses may reveal little about the memory processes mediating performance in implicit and explicit memory tasks (Ostergaard, 1992, p. 419).

We have argued here that Ostergaard's model represents a clear improvement of the technique. The technique of successive tests and contingency analyses is on a firmer footing now. What does seem to need further improvement is the practitioners' understanding of the experiments in which the technique has been used. Because Ostergaard used inappropriate data, he failed to appreciate, and communicate to others, the exact nature of the contribution of his model. We have rectified Ostergaard's mistake and set the record straight with respect to the issue of stochastic independence, at least for the time being. We have also demonstrated the power of Ostergaard's (1992) model in predicting outcomes of experiments in which successive tests are expected to be maximally dependent.

When all this is said and done, it is difficult to suppress the nagging suspicion that correlational methods, including contingency analyses, may indeed be an invention of the devil.

Manuscript received July 1993

REFERENCES

Brown, W. (1923). To what extent is memory measured by a single recall? *Journal of Experimental Psychology*, *6*, 377–382.

Bryant, D.J. (1991). Exceptions to recognition failure as a function of the encoded association between cue and target. *Memory and Cognition*, *19*, 210–219.

Challis, B.H., & Brodbeck, D.R. (1992). Level of processing affects priming in word fragment completion. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 595–607.

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. (1991). The implications of item differences: Commentary on Hintzman and Hartry. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17*, 338–340.

Flexser, A.J., & Tulving, E. (1978). Retrieval independence in recognition and recall. *Psychological Review*, *88*, 327–339.

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

Graf, P., & Ryan, L. (1990). Transfer-appropriate processing for implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 978–992.

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.

Hayman, C.A.G., & Tulving, E. (1989a). Contingent dissociation between recognition and fragment completion: The method of triangulation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 228–240.

Hayman, C.A.G., & Tulving, E. (1989b). Is priming in fragment completion based on a "traceless" memory system? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 941–956.

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

Hintzman, D.L., & Hartry, A.L. (1990). Item effects in recognition and fragment completion: Contingency relations vary for different subsets of words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 955–969.

Jacoby, L.L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior*, *22*, 485–508.

Jacoby, L.L., & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology*, *32*, 300–324.

LeVoi, M.E., Ayton, P.J., Jonckheere, A.R., McClelland, A.G.R., & Rawles, R.E. (1983). Unidimensional memory traces: On the analysis of multiple cued recall. *Journal of Verbal Learning and Verbal Behavior*, *22*, 560–576.

Light, L.L., Singh, A., & Capps, J.L. (1986). Dissociation of memory and awareness in 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*, *38A*, 645–662.

Musen, G., & Treisman, A. (1990). Implicit and explicit memory for visual patterns. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 127–137.

Neely, J.H., & Payne, D.G. (1983). A direct comparison of recognition failure rates for recallable names in episodic semantic memory tests. *Memory and 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., & Gardiner, J.M. (1993). Identifying exceptions in a database of recognition failure studies from 1973 to 1992. *Memory and Cognition*, *21*, 397–410.

Nilsson, L.-G., Law, J., & Tulving, E. (1988). Recognition failure of recallable unique names: Evidence for an empirical law of memory and learning. *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.

Ostergaard, A. (1992). A method for judging measures of stochastic independence: Further comments on the current controversy. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 413–420.

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

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

Roediger, H.L., III, & Blaxton, T.A. (1987). Retrieval modes produce dissociations in memory for surface information. In D.S. Gorfein & R.R. Hoffman (Eds), *Memory and cognitive processes: The Ebbinghaus Centennial Conference*, pp. 349–379. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Roediger, H.L., III, & McDermott, K.B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds), *Handbook of neuropsychology*, Vol. 8, pp. 63–131. Amsterdam: Elsevier.

Roediger, H.L., III, Weldon, M.S., & Challis, B.H. (1989). 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: Essays in honour of Endel Tulving*, pp. 3–42. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Schacter, D.L. (1990). Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. In A. Diamond (Ed.), Development and neural bases of higher cognitive functions. *Annals of the New York Academy of Sciences*, *608*, 543–571.

Schacter, D.L., Cooper, L.A., & Delaney, S.M. (1990). Implicit memory for unfamiliar objects depends on access to structural descriptions. *Journal of Experimental Psychology: General*, *119*, 5–24.

Schacter, D.L., Cooper, L.A., Delaney, S.M., Peterson, M.A., & Tharan, M. (1991). Implicit memory for possible and impossible objects: Constraints on the construction of structural descriptions. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 3–12.

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

Snodgrass, J.G., & Feenan, K. (1990). Priming effects in picture fragment completion: Support for the perceptual closure hypothesis. *Journal of Experimental Psychology: General*, *119*, 276–296.

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

Tulving, E., & Schacter, D.L. (1990). Priming and human memory systems. *Science*, *247*, 301–306.

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.

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*, 352–373.

Tulving, E., Hayman, C.A.G., & Macdonald, C.A. (1991). Long-lasting perceptual priming and semantic learning in amnesia: A case experiment. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17*, 595–617.

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

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

Witherspoon, D., & Moscovitch, M. (1989). Stochastic independence between two implicit memory tests. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 22–30.