THE OXFORD HANDBOOK OF MEMORY

EDITED BY
Endel Tulving
Fergus I. M. Craik

OXFORD
UNIVERSITY PRESS
2000

Contents

4

Contingency Analyses of Memory

MICHAEL J. KAHANA

Human memory subsumes a multitude of processes that undoubtedly rely on many different types of information, operations, and systems. Recognition of the complexity of human memory has driven scholars to broaden the scope of memory tasks that are studied, and to consider not only how a given memory task is performed but also how different kinds of memory tasks are interrelated. It is the study of the similarities and differences among memory tasks that may help us develop predictive models of the underlying information, operations, and systems that support this vital human capacity. The study of human memory thus requires techniques both for dissecting the information-processing components of a single task and for examining the relations between information-processing components of different tasks.

One standard method for separating components both within and between memory tasks is to look for experimental factors that have different effects on different memory tasks, or on different aspects of subjects' performance in a given task. This kind of task analysis has yielded numerous examples of parallel effects and dissociations among tasks (see Richardson-Klavehn & Bjork, 1988, for a review). Converging evidence for dissociations between memory tasks is taken by some researchers as evidence for the operation of multiple memory systems (e.g., Nyberg & Tulving, 1996; Schacter & Tulving, 1994; Tulving, 1985).

Dissociations can also be observed within a single task. For example, dissociations between different portions of the serial position curve (e.g., Glanzer & Cunitz, 1966; Murdock, 1962) have been taken as evidence for the distinction between short-term and long-term memory systems. Techniques for performing task analysis with response time data (Sternberg, 1969) have enabled researchers to distinguish between stages of processing that are relatively independent of one another (see Sternberg, 1998, for a review).

This chapter is concerned with another method of task analysis: contingency analyses applied to the outcomes of successive memory tests. This method has been used to examine the correlation between the outcomes of successive memory tests at the level of an individual subject-item (e.g., DaPolito, 1967; Estes, 1960; Tulving & Wiseman, 1975). The correlation between the observed responses reflects, at least in part, the degree to which their memory processes tap common information, operations, and/or systems. Because the responses are usually dichotomous variables (recall of an item on test 1 and on test 2) these correlations are measured using a 2×2 contingency table. Correlations derived from contingency tables are susceptible to all of the potential confoundings that face correlations between continuous measures. Fears of these potential confoundings have caused many investigators to shy away from using correlational measures to study memory. The position taken in this review is that such avoidance behavior is not well founded, and that these measures can provide information not present in simple measures of memory performance under different conditions.

The chapter is organized into six short sections. The first two sections briefly discuss functional and correlational approaches to provide a context for the discussion of contingency analyses. The fourth presents three applications of the use of contingency analyses in memory research. The next section reviews the controversy over the use of contingency analyses. Finally, the last section offers a summary and some concluding thoughts.

Functional Approaches

In laboratory memory experiments, the basic unit of information is a subject item in a given experimental condition. Under particular retrieval conditions one can ask whether the item is remembered (usually a binary outcome) and how long it takes for the response to be made (response time or RT). In general, RT and accuracy provide complementary but correlated pictures of human behavior (for a discussion of the relation between RT and accuracy in human memory, see Kahana & Loftus, 1999).

Whether or not a target item is remembered depends on (1) the subject, (2) the item, and (3) the experimental condition. In the functional approach, the latter effect is of sole interest. Mean values, computed over many subjects and many items, are compared among conditions and taken as estimates of the "effect" of the manipulated variable. Subject and item differences are typically controlled, either physically or statistically, so that the comparison of performance in one condition with that in another can be made with respect to the independent variable(s) of interest in the experiment. This functional approach teaches us how memory depends on the variables we have manipulated in our experiments.

Comparisons across tasks often reveal interactions. For example, the degree to which a delay between study and test affects retention of a given item is mediated by the kind of memory test employed (e.g., item recognition, fragment completion, free recall). This does not necessarily mean that these different kinds of memory tasks tap different kinds of information. Consider a simple model in which $S(t) = at^{-b}$ defines the strength of each item in memory as a function of time (in this equation, a and b are positive constants). Recognition succeeds when the strength is above a recognition threshold (K_1); recall succeeds when the strength is above a recall threshold (K_2). If $K_1 < K_2$, then there will be a statistical interaction between the retention interval and the kind of test (recognition vs. recall). In this simple model, the interaction results because of the nonlinearity in the forgetting function and not because recognition and recall tap different underlying sources of information.

Correlational Approaches

The correlational approach need not involve manipulating any variables. Rather, one can study the relations between different measures obtained from different tasks, a single task, or even a single response. Consider a single response in a recognition task that asks subjects to make confidence judgments. Murdock and Anderson (1975; see also Koppell, 1977) carefully documented the well-known finding that RT and confidence judgments covary—high-confidence judgments are made faster than low-confidence judgments. This positive correlation is consistent with the view that RT and confidence are both affected by the "strength" of the underlying memory (or, in contemporary terms, the strength of the context-item association).

Within the free-recall task, different aspects of performance are correlated. For example, Tulving (1962, 1966) showed that as subjects learn a list, their recall order becomes increasingly stereotyped (the phenomenon of *subjective organization*). Although plotted as a functional relationship, the experimenter does not directly manipulate either recall probability or subjective organization. Rather, the independent variable, number of study trials, produces correlated effects on both dependent variables.

Examining the correlation between different memory tasks, Kahana and Rizzuto (1999) measured subjects' performance on item-recognition, associative-recognition, and cued-recall tasks. The correlation between item recognition and cued recall was moderate ($r^2 = 0.33$) whereas the correlation between associative recognition and cued recall was high ($r^2 = 0.80$). These subject correlations involve computing an average for each subject across many items. Another correlational technique asks whether the ordering of performance across items in one task or situation is correlated with the ordering of performance across items in another task or situation. If recognition and recall tap common information, we would expect that the correlation across items is positive. For instance, if recognition and recall both tap "strength" but recall requires a higher threshold, we would expect a strong positive correlation across items. Because these techniques involve averaging (across either subjects *or* items), they allow the investigator to compute a Pearson product-moment correlation. Mandler (1959) demonstrated that correlations across subjects and items can often yield very different results. In some cases, correlations across subjects can be positive while correlations across items can be negative. The use of subject and item based correlations is fairly uncommon in memory research (for exceptions, see Rubin, 1981; Underwood, Boruch, & Malmi, 1978).

Contingency Analyses of Successive Tests

Taking this analysis of the recognition-recall relation one step further, we can examine the relation between recognition and recall at the level of individual subject items. To do this, we employ the method of successive tests. In the canonical procedure, subjects study pairs of items (A-B) and are then given two successive tests: item-recognition followed by cued recall. In the item recognition test, the experimenter presents B items from the studied pairs intermixed with nonstudied items (lures). Subjects judge each item as a target or a lure. In the second, cued-recall test, subjects attempt to recall the B items given the A items as cues. The same B items are tested twice—first using a recognition test and then, later, using a recall test. Because we cannot average over subjects or items we compute a *contingency table* from the pairs of outcomes on test 1 and test 2 for each subject item. Table 4.1 gives hypothetical data for successive item-recognition and cued-recall tests. These hypothetical data show the usual advantage for recognition over recall. However, examining the contingency table allows us to say something stronger. For example, according to a simple strength theory of memory, recognition and recall tap exactly the same information, but recall requires a higher threshold. Strictly speaking, this means that an item that is recalled will always be recognized, making the correlation between recognition and recall, at the level of subject items, exactly one. The hypothetical data in table 4.1, and the real data of Tulving and Thomson (1973), indicate that there is recognition failure of recallable items —contradicting the predictions of strength theory.

Because the test outcomes are binary variables, we compute the subject-item correlation between recognition and recall by tabulating the data in a contingency table. This correlation is often referred to as contingency or dependency, as it is calculated from the cells in

Table 4.1 Hypothetical data comparing recognition and recall performance for the same subject-items tested successively.

Subject-Item	Test 1 (recognition)	Test 2 (recall)
1	1	0
2	0	0
3	1	1
4	1	0
5	0	0
6	1	0
7	1	1
8	0	1
9	1	1
10	1	0
11	0	0
12	1	0
13	1	1
14	1	0
15	1	1
16	1	0
17	1	0
18	1	1
19	0	0
20	1	0
21	0	1
22	1	1
23	0	0
24	1	1
25	0	0
26	1	1
Mean	0.69	0.42

a 2 × 2 contingency table. Yule's Q is a popular measure of correlation for 2 × 2 contingency tables (Bishop, Feinberg, & Holland, 1975). Like a standard Pearson correlation, Yule's Q varies from −1.0 (perfect negative correlation) to +1.0 (perfect positive correlation). For a discussion of other correlational measures that are applicable to contingency tables, see Poldrack (1996).

Table 4.2 shows the results of a contingency analysis applied to the hypothetical data in table 4.1. For each subject item, the combination of outcomes on test 1 and test 2 determine the values of the four cells in the contingency table (A, B, C, & D). Yule's Q is given by the equation:

$$Q=\frac{A\times D-B\times C}{A\times D+B\times C}.$$

For the hypothetical data shown in table 4.1, the dependency between the item recognition and cued recall, as measured by Yule's Q, is 0.5. This dependency of 0.5, measured at the level of subject items, does not mean that the correlation across subjects is 0.5 or that the correlation across items is 0.5. As pointed out by Tulving (1985), correlations at the level of subjects, items, and subject items can yield very different values.

This example was chosen because there is a vast literature examining the relation between recognition and recall using the method of successive tests. This literature, first surveyed by Tulving and Wiseman (1975) and more recently reviewed by Nilsson and Gardiner (1993), reveals an invariance: successive-item recognition and cued-recall tasks almost always yield moderate correlations, with Yule's Q rarely deviating from the range 0.3 to 0.75. This makes the successive-testing data inconsistent with a simple strength-threshold theory (Tulving, 1983), and also with certain distributed memory models (Kahana & Rizzuto, 1999).

Table 4.2 Contingency table for the hypothetical data shown in table 1.

		Test 1: 1	Test 1: 0
Test 2	1	a = 9	b = 2
	0	c = 9	d = 6

Together with Yule's Q, the percent correct for tests 1 and 2 fully characterize the data in our 2 × 2 contingency table. Without the correlational information, the accuracy data would do little to constrain theory. Because theories of memory make claims about the processes acting on a given subject item, it is important to consider not just accuracy data but also the correlation between tests.

The foregoing example illustrates the comparison of tasks using contingencies of outcomes on successive tests. In comparing performance across the two tests, it is important to recognize that the successive tests do not necessarily measure the "same thing." The first measure is of the consequences of study—subject to all the input/output interference effects that may operate on retrieval of individual list items (e.g., Tulving & Arbuckle, 1966). The second measure is affected not only by study and the interpolated conditions but also by the earlier test and its outcome.

One can also analyze contingencies across successive trials of the same task. This approach was fruitfully employed by Estes (1960) in his studies of one-trial learning and by Tulving (1964) in his analysis of inter- vs. intra-trial forgetting in multitrial free recall.

Empirical Regularities in Successive Testing Experiments

There are many interesting examples of the use of contingency analyses, but the most well-studied problems include comparisons of successive episodic memory tasks (Nilsson & Gardiner, 1993; Tulving & Wiseman, 1975), successive explicit and implicit memory tasks (Hayman & Tulving, 1989a, 1989b; Tulving, Schacter, & Stark, 1982), and successive implicit memory tasks (e.g., Hayman & Tulving, 1989b; Tulving & Hayman, 1995; Witherspoon & Moscovitch, 1989). Other important applications of successive tests include demonstrations of one-trial learning (e.g., Estes, 1960) and demonstrations of the independence of A-B and A-C associations in the retroaction (e.g., Greeno, James, & Dapolito, 1971). In this section, three applications are discussed: independence of A-B and A-C associations, the recognition failure paradigm, and some recent results pertaining to the classic question of associative symmetry versus independent associations. These three examples were chosen because they yield correlations of approximately 0.1, 0.5, and 0.9, respectively.

Analysis of Competing A-B and A-C Associations

The classic associative interference theory of verbal learning played a prominent role in guiding a generation of research in verbal learning (see Postman & Underwood, 1973, for a review). A major technique used to study associative interference is the A-B/A-C paradigm. Subjects first master a list of A-B pairs. Next, subjects study a second list of A-C pairs to some degree of mastery. Briggs (1954) showed that with increasing trials of A-C learning comes a decrease in subjects' ability to recall the original A-B associations. To examine this retroactive interference effect under conditions designed to minimize response competition, Barnes and Underwood (1959) proposed the now-classic MMFR procedure (MMFR stands for modified-modified free recall). In this procedure, subjects study the A-B list, then study the A-C list, and then are given each of the A items and asked to recall *both* the B and C responses *in any order.* With increasing trials of A-C learning, subjects recall fewer B items and more C items, demonstrating the classic retroactive interference phenomenon. According to the Melton-Underwood unlearning-recovery hypothesis—a central tenet of associative interference theory—the decrease in A-B recall following A-C learning results from *specific* unlearning of the individual A-B associations. This contrasts with the earlier view (McGeoch, 1942) that associations are independent, but that the stronger response dominates. Here we have two very different theories that both explain the basic data on retroactive interference.

Applying a contingency analysis to the MMFR data, Dapolito (1967) examined the dependence between B and C recall. According to the unlearning-recovery hypothesis, learning A_i-C_i causes specific weakening of the A_i-B_i association. As a consequence, recall of B_i and C_i should be negatively correlated. In contrast to these predictions, Dapolito found near independence between recall of B_i and C_i. Numerous additional studies supported the independence position (Abra, 1969; Greeno et al., 1971; Martin, 1971; Wichawut & Martin, 1971). Because these studies did not report their results in terms of Yule's Q, it is hard to compare their observations with other studies of correlations among successive tests. Appendix A presents a database of 32 experimental conditions obtained using the MMFR technique. Across these conditions, the mean value of Yule's Q is 0.08, with a standard deviation of 0.31. Again, contingency analyses reveal a regularity of human memory that distinguishes among theories of associative interference. Several major mathematical models of human memory (e.g., Murdock, 1982, 1997; Chappell & Humphreys, 1994; Mensink & Raaijmakers, 1988) have been framed in a way that captures the basic phenomenon of A-B, A-C independence.

The Recognition Failure Paradigm (Tulving & Thomson, 1973)

Perhaps the most common application of contingency analyses to the study of memory is the influential, if somewhat controversial, recognition-failure paradigm (Flexser & Tulving, 1978; Tulving & Thomson, 1973; Tulving & Wiseman, 1975). In this approach, discussed briefly in the previous section, subjects study a list of word pairs and are then tested successively, first by item recognition and then by cued recall. Tulving and Wiseman (1975) observed that the dependency relation between item recognition and cued recall, across a wide range of experimental conditions, was well fit by a quadratic formula relating the conditional probability of recognition given recall to the probability of recognition itself. This function, known as the Tulving-Wiseman function, describes a moderate degree of dependency between item recognition and cued recall.

Analyses based on the conditional probability of recognition given recall, P(Rn|Rc), are subject to a number of problems: First, P(Rn|Rc) is constrained to be less than P(Rn)/P(Rc) in experiments where the probability of recall exceeds that of recognition (Hintzman & Hartry, 1990; Hintzman, 1992). Second, the predicted values of Yule's Q, derived from the Tulving-Wiseman function, can be less than −1 or greater than +1 for some combinations of P(Rn) and P(Rc). These potential constraints can be avoided by examining results in terms of both Yule's Q and the probability of success on test 1 and on test 2. The cells in a 2 × 2 contingency table are completely determined by either (1) knowing the probabilities of successful outcomes on tests one and two as well as the correlation (Yule's Q) between these outcomes, or (2) knowing the probability of success on test 1, the probability of success on test 2 conditional upon test 1 success, and the probability of success on test 2 conditional upon test 1 failure. (A proof of this assertion is available from the author upon request.) This later approach is used by Humphreys and Bowyer (1980) as well as Batchelder and Riefer (1995).

Using a measure of dependency such as Yule's Q, one typically obtains a value of approximately 0.55 in these experiments. Not all experiments using the successive testing methodology yield the moderate dependencies described by the Tulving-Wiseman function (Nilsson, Law, & Tulving, 1988; Nilsson & Gardiner, 1993). Higher recognition-recall dependencies can result from shallow encoding or from semantic redundancy of study pairs. Nilsson and Gardiner (1991) refer to these cases as boundary conditions on the Tulving-Wiseman function. A great many studies have looked at the dependency between item recognition and cued recall under varying conditions (see Nilsson & Gardiner, 1993, for a partial review). Examining the correlations summarized by Nilsson and Gardiner (1993) reveals significant variation around the mean value of 0.55. The reason for this variation is that many of these studies were especially constructed to push the correlation up or down. Appendix B presents a database of Yule's Q values from studies that gathered significant amounts of data using standard methods (i.e., subjects study a list of common word pairs and are then given a yes/no recognition test followed by a cued-recall test). Across these studies, the mean value of Yule's Q is 0.55 and the standard deviation is 0.12. The consistently obtained moderate correlation between item recognition and cued recall represents a basic fact of human memory.

Kahana and Rizzuto (1999) have shown that the moderate dependency obtained in successive item recognition and cued recall tasks is also found in successive item and associative recognition tasks (Q = 0.59). This suggests that the reason for the moderate dependency is that the one test taps item and the other test taps associative information. In contrast, when both tasks tap associative information (e.g., successive *associative* recognition and cued recall), the correlation rises substantially (Q = 0.81). Finally, for identical associative recognition tasks correlations approach unity (Q = 0.94).

Though the data are clear, there is still much debate over the interpretation of these findings. Metcalfe (1991) shows that under certain conditions, CHARM (a mathematical memory model that assumes composite and distributed storage of auto- and hetero-associative information) can account for both the Tulving-Wiseman function and conditions that result in deviations from the function. Kahana and Rizzuto (1999) found that several classes of memory models including Metcalfe's (1982, 1985) CHARM model, Murdock's (1982) TODAM model and the matrix model of Humphreys, Bain, and Pike (1989) can all produce the moderate dependencies required, but only if you allow for variability in the goodness of encoding (see also Hintzman, 1987). Each of the preceding models assumes that item and associative information have distinct representations. Models that assume identical representations for individual items and associations produce correlations that are too high, deviating from the experimental data for all reasonable parameter values. These results show that if models make explicit predictions about subject performance in the successive testing paradigm, contingency analyses place constraints on the models and, in doing so, provide insight into the function of human memory.

Associative Symmetry vs. Independent Associations

Another basic question in the study of human memory is the nature of associations. Two models of association are present in the classical literature: The independent associations model considers associations to be unidirectional links between stored items (e.g., Ebbinghaus, 1885/1913; Robinson, 1932). The holistic model considers associations to be newly formed patterns combining elements of each stored item (Kohler, 1940). In this model, forward and backward associations between items are symmetrical (Asch & Ebenholtz, 1962). Early studies addressing the differences between these two positions focused on whether forward recall is easier than backward recall (see Ekstrand, 1966, and Kahana, 1999, for a review). The evidence, from numerous studies, suggests that order of study has a minimal effect on associative strength. Findings of equivalent forward and backward retrieval have been taken as evidence for the position of associative symmetry; findings of asymmetric retrieval have been taken as evidence for the position of independent associations (e.g., Wollen, 1970a,b; Wolford, 1971).

A much stronger test of associative symmetry requires that for a given studied pair (A-B), the strength of the forward and backward associations, and hence the recall probabilities, must be identical. Evidence relevant to this question may be gleaned from contingency analyses applied to the outcomes of successive cued-recall tests. First, one association is tested (say, A is presented as a cue to recall B) and then the other association is tested (B is presented as a cue to recall A).

Kahana (1999) had subjects study lists of word pairs and then gave successive cued-recall tests of all pairs in the list. Across the two successive tests, word pairs were tested in all combinations of forward and backward orders: T1-forward/T2-forward, T1-forward/T2-backward, T1-backward/T2-forward, T1-backward/T2-backward. To test the symmetry principle, Yule's Q was computed for identical and reverse successive tests (separate Yule's Q values were calculated for each subject). For identical successive tests (i.e., forward-forward or backward-backward), the average Yule's Q value was 0.88 (SE = 0.012). For reverse successive tests (i.e., forward-backward or backward-forward), the average Yule's Q value was 0.91 (SE = 0.017). These effects did not differ statistically. The independent associations model of paired-associate learning (Wolford, 1971; Wollen, 1970a,b) can account for asymmetries that pose a resolvable challenge to symmetrical associative models (see Kahana, 1999), but they cannot explain the finding that correlations between forward and backward recall are near unity.

Other Applications

The previous three applications illustrated how intertask contingencies can vary quite dramatically and how they can be used to constrain theories of memory. There are a number of other important applications of contingency analyses that are not reviewed here. In particular, contingency analyses have been extensively applied to the study of implicit memory (Hayman & Tulving, 1989a, 1989b; Tulving et al., 1982; Tulving, Hayman, & Macdonald, 1991; Tulving & Hayman, 1995; Witherspoon & Moscovitch, 1989). These studies have shown that successive implicit tasks can yield dependencies ranging from near zero (with nonoverlapping cues) to around 0.5 (with identical cues). In contrast, successive explicit tasks yield dependencies ranging from around 0.5 (with different cues) to 1.0 (with identical cues). An important aspect of these results is that the dependency between fragment completion tasks can be dramatically affected by subjects' intentionality (i.e., whether or not subjects are instructed to focus retrieval on their memory for the study list). Tulving and colleagues have used this evidence, together with findings of functional independence, to support a multiple memory systems view (e.g., Schacter & Tulving, 1994).

The Controversy

Despite the fruitful application of contingency analyses to a broad range of memory paradigms, there has been considerable controversy surrounding their use as an analytic tool. Hintzman, a vocal critic, has argued that correlations between tasks, as measured by Yule's Q, are at best difficult to interpret and at worst uninterpretable. His arguments are based on Simpson's paradox (Hintzman, 1980). This refers to the fact that collapsing data across subjects or items *can* give rise to relations that were not present in the pre-collapsed data. In practice, what does this mean? Here we examine two cases, both using our familiar example of successive item-recognition and cued-recall tests (e.g., Tulving & Thomson, 1973). Recall that in these experiments, Yule's Q is almost always between 0.30 and 0.75.

Case 1. The experimenter presents a long list of A_i-B_i word pairs for study. As the list is being presented, subjects adopt the following strategy: they attend to the first few pairs and then close their eyes and keep rehearsing those pairs throughout the duration of the list presentation. Upon tabulating our results, we discover that all responses are segregated into cells A (+/+) and D (–/–) of the contingency table, with D (–/–) having the majority of responses. The few items that subjects rehearsed were both recognized on test 1 and recalled on test 2 (accounting for the responses in cell A). The remaining, unseen items, were neither recognized nor recalled (accounting for the responses in cell D). The resulting correlation is exactly +1.0. This result does not reflect common operations underlying recognition and recall; rather it is a spurious correlation induced by a third actor—variability in encoding.

Case 2. Here again, the experimenter presents a list of A_i-B_i pairs, and then gives successive item recognition and cued recall tests for B_i. Suppose we now divide our pairs into two groups: in group 1, we make the B-items easy to recognize but hard to recall, and in group 2, we make the B-items hard to recognize and easy to recall. We can manipulate recognition of B_i by varying its word frequency or its similarity to the lures. We can manipulate recall of B_i by varying the associative relations between A_i-B_i and between A_i-B_j. After tabulating our contingency table, most of the pairs in group 1 would be in cell C (+/–) and most of the pairs in group 2 would be in cell B (–/+). In this case, the correlation between item recognition and cued recall would be close to –1.0. This illustrates another spurious correlation: we identified factors that have different effects on recognition and recall and then specially selected pairs to induce a negative correlation.

In an attempt to show Simpson's paradox at work in successive recognition and fragment completion tests, Hintzman and Hartry (1990) found that selecting different subsets of items could produce large changes in the observed correlation between tasks. Such demonstrations do not teach us very much. They do point out, as all practitioners who use correlations ought to know, that spurious correlations can occur. In particular, variability that affects the two outcome variables in a correlated manner can induce an increase or decrease in the observed correlation.

Suppose that variability in subject ability and item difficulty both have positive effects on memory performance. In this case, the observed value of Yule's Q would be somewhat higher when collapsing across subjects or items. Flexser (1981) provided a useful technique for adjusting Yule's Q to account for these sources of variability. Another approach requires the collection of sufficient data across subjects and items to permit separate Yule's Q analyses for contingency tables that are collapsed across only subjects, or only items. When these techniques have been used, the correlation between successive tasks has *not* changed dramatically.

One problem with these approaches is that they do not adjust for potential variability at the level of subject items *that is caused by factors not intrinsic to the tasks being studied* (e.g., trial by trial fluctuations in attention). If this external source of variability, affecting the encoding of information important for both test 1 and test 2, is *much* larger than the measurable variability across subjects and items, then it is possible that the "true" correlation is smaller than what is observed experimentally.

One way to address this potential problem is to create lists in which the variability is artificially increased by mixing strong and weak items (where strength is manipulated by varying number of repetitions or presentation rate). One can assess the effect of variability on the dependency relations among tasks by comparing these mixed lists with pure lists (in which presentation rate, or number of repetitions, is uniform across all study pairs). Kahana and Rizzuto (1999) conducted two experiments of this kind and found significant increases in Yule's Q for the specially constructed, high-variability lists. However, the effects were extremely small (the largest changes in variability only increased Yule's Q by about 0.10). In contrast, informational manipulations produced dramatic changes in the correlations between successive tests. These findings strongly suggest that variability in goodness-of-encoding only plays a minor role in determining the correlations between successive memory tasks.

A final class of criticisms relate to the "priming" of second test performance by the first test (e.g., Humphreys & Bowyer, 1980). As stated previously, it is well known that the first test may affect performance on the second test, and therefore the second test is no longer a direct measure of the study event. Rather, the second test reflects both the study event and the effect of the first test on memory. Humphreys and Bowyer argue that the dependency between recognition and recall is increased by differential facilitation of later recall by prior recognition. In essence, recognized items may be strengthened during the recognition test more than nonrecognized items. This transfer effect enhances cued-recall performance for the recognized items more than for the nonrecognized items, thereby increasing the dependency between item recognition and cued recall. In several experiments, Humphreys and Bowyer found higher cued-recall performance for those items that were tested in the recognition phase (as compared with items only tested in the cued-recall phase).

If one assumes that only recognized items prime subsequent recall, Humphreys and Bowyer show that the observed moderate correlation between recognition and recall could be largely due to this priming effect. The problem with this account is that nonrecognized items can also facilitate subsequent recall (Begg, 1979; Donnelley, 1988). In addition, even experiments that fail to show significant "priming" yield moderate dependency between item recognition and cued recall (Wiseman & Tulving, 1976). These results suggest that although the outcome of the first recognition test does influence subsequent recall, this effect accounts only for some part of the observed dependency between these tasks.

Where, then, do we stand in the face of these potential confounds? Rather than running away from the complexities of contingency analyses we need to proceed with caution. Researchers have managed to obtain highly reliable and replicable results using the method of successive tests, and the variation in dependencies among tasks have been consistent and theoretically interpretable (Martin, 1981; Gardiner, 1991).

Conclusions

This chapter reviewed the use of contingency analyses applied to successive memory tasks. As shown in the applications section, the correlation between successive tests varies in reliable ways across different task comparisons. Successive item-recognition and cued-recall tasks yield moderate levels of dependency (Yule's $Q \approx 0.55$) whereas successive tests of episodic memory with identical cues or cues containing identical information yield very high dependencies (Yule's $Q \approx 0.90$).With implicit memory tasks, a very different pattern is observed: Yule's Q ranges from 0.1 to 0.5 as a function of cue overlap. This difference between implicit and explicit memory tasks has provided evidence supporting the multiple memory systems view (e.g., Tulving, 1985; Schacter & Tulving, 1994). Although other explanations may well exist, contingency analyses have played an important role in this debate.

Applying contingency analyses to the relation between recognition and recall, researchers have shown that experimental variables that have a significant effect on overall levels of performance do not seem to affect the task-task contingencies (see Flexser & Tulving, 1978; Kahana & Rizzuto, 1999; Nilsson & Gardiner, 1993; Wiseman & Tulving, 1976). This finding highlights an important feature of intertask contingencies: they represent an added dependent variable that is independent of overall performance levels on the two successive tasks.

These and other examples of contingency analyses, reviewed in this chapter and elsewhere (e.g., Tulving, 1985; Tulving & Hayman, 1995) illustrate a striking feature of the correlations between successive tests: the dependency varies systematically with the nature of the two tests. Even in a situation in which the first test remains the same, the study conditions are the same, and the target items are the same, the exact form of the second test can dramatically alter the observed correlation (see Hayman & Tulving, 1989a; Tulving et al., 1991).

Repeated testing of the knowledge that subjects acquire in the laboratory has been of immense value in enriching our understanding of how human memory works—this despite the fear that many students of memory have exhibited, and problems of interpretation that require a different approach than that adopted for the study of the effects of independent variables. In this respect, these experiments resemble life in which, too, the results of an experience are frequently tested repeatedly.

Appendix A

A selected database of studies that examined the correlation between B and C responses in the MMFR paradigm. The mean Yule's Q value for the included conditions was 0.08 ± .11 (95% confidence interval).

Study	Condition	Yule's Q
Abra (1969)	48 hr.–0 hr.	−0.22
Abra (1969)	48 hr.–24 hr.	−0.13
Abra (1969)	48 hr.–48 hr.	0.47
Abra (1969)	24 hr.–24 hr.	−0.10
Delprato (1972)	AC(4,2)	−0.05
Delprato (1972)	AC(4)	0.29
Delprato (1972)	AC(8,4)	−0.32
Delprato (1972)	AC(8,8)	−0.16
Delprato (1972)	AC(16,8)	−0.38
Delprato (1972)	AC(16,16)	0.14
Koppenaal (1978)	Retention Interval = 1 min.	−0.32
Koppenaal (1978)	Retention Interval = 20 min.	0.00
Koppenaal (1978)	Retention Interval = 90 min.	0.37
Koppenaal (1978)	Retention Interval = 6 hours	−0.62
Koppenaal (1978)	Retention Interval = 24 hours	0.00
Koppenaal (1978)	Retention Interval = 72 hours	0.06
Koppenaal (1978)	Retention Interval = 1 week	0.23
Postman (1964)	set 1	−0.32
Postman (1964)	set 2	0.31
Postman (1964)	set 3	0.35
Postman and Gray (1977)	Substitution	−0.13
Tulving and Watkins (1974)	BC	0.16
Tulving and Watkins (1974)	*B′C′*	0.44
Tulving and Watkins (1974)	*B′C*	0.51
Tulving and Watkins (1974)	*BC′*	0.53
Tulving and Watkins (1974)	*BC*	0.39
Tulving and Watkins (1974)	*B′C′*	0.48
Tulving and Watkins (1974)	*B′C*	0.07
Tulving and Watkins (1974)	*BC′*	0.62
Wichawut and Martin (1971)	AC(4)	−0.24
Wichawut and Martin (1971)	AC(8)	0.13
Wichawut and Martin (1971)	AC(12)	0.00

Appendix B

A selected database of successive item recognition—cued-recall experimental conditions. Studies that used pairs of common English nouns, or weak associates, were included; studies that used strong associates or non-English materials were excluded. Studies that used categorization or free association before the successive tests were excluded. Finally, experiments where the standard error on Yule's Q was greater than 0.2 were also omitted. The mean Yule's Q value for included conditions was 0.55 ± 0.05 (95% confidence interval).

Study	Condition	Yule's Q
Bartling and Thompson (1977)	Noun-noun condition	0.50
Wallace (1978)	Weak cue target pairs, Rn, cued Rc	0.43
Wallace (1978)	Weak cue target pairs, Rn, cued Rc	0.76
Wallace (1978)	Weak cue target pairs, Rn, cued Rc, no lures	0.64
Begg (1979)	Meaningful instructions	0.19
Begg (1979)	Rote instructions	0.63
Bowyer & Humphreys (1979)	Weak cue-target pairs, between subject priming	0.41
Bowyer & Humphreys (1979)	Weak cue target pairs, within subject priming	0.53
Fisher (1979)	Repetition rehearsal, high-low cue target assoc.	0.75
Vining and Nelson (1979)	Weak cue target pairs	0.59
Vining and Nelson (1979)	Unrelated cue target pairs	0.51
Gardiner (1988)	Cued recall overall	0.66
Gardiner (1988)	Cued recall (ctrl), strict	0.65
Gardiner (1988)	Cued recall overall, strict	0.54
Gardiner (1988)	Cued recall (ctrl), lenient	0.73
Gardiner (1988)	Cued recall overall, lenient	0.49
Sandberg (1988)	Weak cue target pairs, cued recall instruction	0.55
Kahana and Rizzuto (1999)	Experiment 1 (Pure Strong)	0.54
Kahana and Rizzuto (1999)	Experiment 1 (Pure Weak)	0.49
Kahana and Rizzuto (1999)	Experiment 2 (Pure Strong)	0.49
Kahana and Rizzuto (1999)	Experiment 2 (Pure Weak)	0.52
Kahana and Rizzuto (1999)	Experiment 3 (item recognition)	0.57

Acknowledgments This research was funded by N.I.H. grant MH55687 to Brandeis University. I am grateful to Ben Murdock for his comments on a previous version of this chapter. Thanks are also due to Dan Rizzutto and Marc Howard for very helpful comments and discussion. Correspondence should be addressed to M. J. Kahana, Center for Complex Systems, Brandeis University, Waltham, MA, 02454-9110. E-mail may be sent to: kahana@cs.brandeis.edu.

References

Abra, J. C. (1969). List-1 unlearning and recovery as a function of the point of interpolated learning. *Journal of Verbal Learning and Verbal Behavior, 8*, 494–500.

Asch, S. E., & Ebenholtz, S. M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society, 106*, 135–163.

Barnes, J. M., & Underwood, B. J. (1959). Fate of first-list associations in transfer theory. *Journal of Experimental Psychology, 58*, 97–105.

Bartling, C. A., & Thompson, P. (1977). Encoding specificity: Retrieval asymmetry in the recognition failure paradigm. *Journal of Experimental Psychology: Human Learning and Memory, 3(6)*, 690–700.

Batchelder, W. H., & Riefer, D. M. (1995). A multinomial modeling analysis of the recognition-failure paradigm. *Memory & Cognition, 23*, 611–630.

Begg, I. (1979). Trace loss and the recognition failure of unrecalled words. *Memory & Cognition, 7*, 113–123.

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

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

Briggs, G. E. (1954). Acquisition, extinction, and recovery functions in retroactive inhibition. *Journal of Experimental Psychology, 47*, 285–293.

Chappell, M., & Humphreys, M. (1994). An autoassociative neural network for sparse representations: Analysis and application to models of recognition and cued recall. *Psychological Review, 101*, 103–128.

Dapolito, F. J. (1967). Proactive effects with independent retrieval of competing responses. *Dissertation Abstracts International, 27*, 2522–2523.

Delprato, D. (1972). Pair specific effects in retroactive inhibition. *Journal of Verbal Learning and Verbal Behavior, 11*, 566–572.

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

Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology.* New York: Teachers College, Columbia University.

Ekstrand, B. R. (1966). Backward associations. *Psychological Bulletin, 65*, 50–64.

Estes, W. K. (1960). Learning theory and the new "mental chemistry." *Psychological Review, 67*, 207–223.

Fisher, R. P. (1979). Retrieval operations in cued recall and recognition. *Memory & Cognition, 7*, 224–231.

Flexser, A. (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.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*, 309–313.

Gardiner, J. M. (1991). Contingency relations in successive tests: Accidents do not happen. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17*, 334–337.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5*, 351–360.

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.

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

Hayman, G. A. C., & Tulving, E. (1989b). Is priming in fragment completion based on a "traceless" memory system? *Journal of Experimental Psychology: Learning, Memory, & 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. (1987). Recognition and recall in minerva 2: Analysis of the 'recognition-failure' paradigm. In P. Morris (Ed.), *Modelling cognition* (pp. 215–229). New York: Wiley.

Hintzman, D. L. (1992). Mathematical constraints on the Tulving-Wiseman function. *Psychological Review, 99*, 536–542.

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, & Cognition, 16*, 965–969.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review, 96*, 208–233.

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

Kahana, M. J. (1999). Associative symmetry and memory theory. Submitted for publication.

Kahana, M. J., & Loftus, G. (1999). Response time versus accuracy in human memory. In R. J. Sternberg (Ed.), *The Nature of Cognition* (pp. 323–384). Cambridge, CA: MIT Press.

Kahana, M. J., & Rizzuto, D. S. (1999). An analysis of the recognition-recall relation in four distributed memory models. *Manuscript submitted for publication.*

Köhler, W. (1940). *Dynamics in Psychology.* New York: Liveright.

Koppell, S. (1977). Decision latencies in recognition memory: A signal detection theory analysis. *Journal of Experimental Psychology: Human Learning & Memory, 3*, 445–457.

Koppenaal, R. J. (1963). Time changes in the strengths of a-b, a-c lists; spontaneous recovery? *Journal of Verbal Learning and Verbal Behavior, 2*, 310–319.

Mandler, G. (1959). Stimulus variables and subject variables: A caution. *Psychological Review, 66*, 145–149.

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.

McGeoch, J. A. (1942). *The psychology of human learning: An introduction.* New York: Longmans.

Mensink, G. J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95*, 434–455.

Metcalfe, J. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review, 92*, 1–38.

Metcalfe, J. (1991). Recognition failure and the composite memory trace in CHARM. *Psychological Review, 98*, 529–553.

Metcalfe-Eich, J. (1982). A composite holographic associative recall model. *Psychological Review, 89*, 627–661.

Murdock, B. B. (1962). The serial position curve in free recall. *Journal of Experimental Psychology, 64*, 482–488.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89*, 609–626.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104*, 839–862.

Murdock, B. B., & Anderson, R. E. (1975). Encoding, storage and retrieval of item information. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 145–194). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Nilsson, L. G., & Gardiner, J. M. (1991). Memory theory and the boundary conditions of the Tulving-Wiseman law. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 57–74). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nilsson, L. G., & Gardiner, J. M. (1993). Identifying exceptions in a database of recognition failure studies from 1973 to 1992. *Memory & 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.

Nyberg, L., & Tulving, E. (1996). Classifying human long-term memory: Evidence from converging dissociations. *European Journal of Cognitive Psychology, 8*, 163–183.

Poldrack, R. A. (1996). On testing for stochastic dissociations. *Psychonomic Bulletin and Review, 3*, 434–448.

Postman, L. (1964). Studies of learning to learn ii. changes in transfer as a function of practice. *Journal of Verbal Learning and Verbal Behavior, 3*, 437–447.

Postman, L., & Gray, W. (1977). Maintenance of prior associations and proactive inhibition. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 255–263.

Postman, L., & Underwood, B. (1973). Critical issues in interference theory. *Memory & Cognition, 1*, 19–40.

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

Robinson, E. S. (1932). *Association theory today; an essay in systematic psychology.* New York: The Century Co.

Rubin, D. C. (1981). First-order approximation to english, second-order approximation to english, and orthographic neighbor ratio norms for 925 nouns. *Behavior Research Methods, Instruments and Computers, 13*, 713–721.

Schacter, D., & Tulving, E. (1994). *Memory Systems 1994.* Cambridge, MA: MIT Press.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30*, 276–315.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *Invitation to Cognitive Science: Vol. 4.: Methods, Models, and Conceptual Issues 2nd ed.* (pp. 703–861).

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review, 69*, 344–354.

Tulving, E. (1964). Intra-trial and inter-trial retention: Notes towards a theory of free recall verbal learning. *Psychological Review, 71*, 219–237.

Tulving, E. (1966). Subjective organization and effects of repetition in multi-trial free-recall learning. *Journal of Experimental Psychology, 72*, 145–150.

Tulving, E. (1983). *Elements of Episodic Memory.* New York: Oxford.

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

Tulving, E., & Arbuckle, T. Y. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology, 72*, 145–150.

Tulving, E., Hayman, C. A. G., & Macdonald, C. A. (1991). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Verbal Learning and Verbal Behavior, 17*, 595–617.

Tulving, E., & Hayman, C. A. G. (1995). On measurement of priming: What is the correct baseline? *European Journal of Cognitive Psychology, 7*, 13–18.

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. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior, 13*, 181–193.

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

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General, 107*, 393–419.

Vining, S. K., & Nelson, T. O. (1979). Some constraints on the generality and interpretation of the recognition failure of recallable words. *American Journal of Psychology, 92*, 257–276.

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

Wichawut, C., & Martin, E. (1971). Independence of a–b and a–c associations in retroaction. *Journal of Verbal Learning and Verbal Behavior, 10*, 316–321.

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

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

Wolford, G. (1971). Function of distinct associations for paired-associate performance. *Psychological Review, 78*, 303–313.

Wollen, K. (1970a). Effects of instructional set and materials upon forward and backward learning. *Journal of Experimental Psychology, 85*, 275–277.

Wollen, K. (1970b). Effects of set to learn A-B or B-A upon A-B and B-A tests. *Journal of Experimental Psychology, 86*, 186–189.

Part II: Memory in the Laboratory