Psychological Review
1993, Vol. 100, No. 1, 149–153

0033-295X/93/$3.00

Recognition-Failure Constraints and the Average Maximum

Arthur J. Flexser and Endel Tulving

The Tulving-Wiseman function of recognition failure of recallable words is defined by the relation between the conditional probability of recognition given recall, P(Rn|Rc), and the probability of recognition, P(Rn). Hintzman (1993) proposed a distinction between algebraic maximum and average maximum values of P(Rn|Rc) to support his earlier claims (Hintzman, 1992) that the Tulving-Wiseman function is a mathematical artifact. This article shows that the distinction depends on a crucial assumption of Hintzman's argument that is unjustified on both empirical and rational grounds. Under a more reasonable assumption, there is no difference between the 2 maxima, so that the reality of the Tulving-Wiseman function remains unchanged.

Hintzman (1992) claimed that the Tulving-Wiseman function (Flexser & Tulving, 1978; Tulving & Wiseman, 1975) is an artifact. His argument was that the regularity that appears in data from recognition-failure experiments is a consequence of mathematical constraints that restrict the upper bound of the conditional probability of recognition given recall, P(Rn|Rc). Tulving and Flexser (1992) showed that Hintzman's rational argument was at variance with empirical data. Their demonstration was based on the fact that mathematical constraints of the kind discussed by Hintzman do not apply in cases where recognition is at least as high as recall; they apply only in cases where recall is greater than recognition. In the former situation, there are no mathematical constraints: The mathematical upper bound of P(Rn|Rc) is 1. In the latter situation, constraints are present: The upper bound of P(Rn|Rc) is P(Rn)/P(Rc).

Tulving and Flexser (1992) presented a quantitative comparison of the data from a large number of experiments with and without mathematical constraints and showed that Hintzman's (1992) claim was invalid. The location of the data points on the recognition-failure plots for the constrained and unconstrained experimental conditions was substantially the same, with both sets of points in close proximity to the Tulving-Wiseman function. Tulving and Flexser therefore concluded that mathematical constraints cannot account for the adherence of recognition-failure data to the Tulving-Wiseman function.

Hintzman (1993) claims that Tulving and Flexser's (1992) empirical demonstration that his original argument was not valid is in turn not valid because it was based on a false dichotomy between mathematically constrained (*ratio-constrained*) and unconstrained (*unity-constrained*) conditions. At the heart of this new argument is the claim that the dichotomy depends on the assumption of zero variability in recognition and recall within experimental conditions. Hintzman (1993) argued, using simulated data, that if the presence of variability is taken into consideration, P(Rn|Rc) is constrained to be less than unity even when recognition exceeds recall for an experimental condition.

In this article, we demonstrate that Hintzman's latest criticism of the Tulving-Wiseman function is based on an unjustified assumption that is implicit in his simulation procedure, an assumption that is at variance with empirical facts.

The Average Maximum

To demonstrate the claimed falsity of Tulving and Flexser's (1992) dichotomization, Hintzman introduced a concept he referred to as the *average maximum,* as distinguished from the maximum value of P(Rn|Rc) that is imposed by purely mathematical constraints, which he called the *algebraic maximum.* It is the average maximum that Hintzman (1993, p. 147) regarded as the true or actual upper bound on P(Rn|Rc), constituting a "variance-induced ceiling" (p. 144); the algebraic maximum was considered to act as a "kind of repulsive barrier" (p. 145). Only when the recognition and recall variances are zero, Hintzman asserted, do the average and algebraic maxima coincide; otherwise, he claimed, the average maximum imposes a more stringent upper bound on P(Rn|Rc) than that imposed by the algebraic maximum.

For purposes of deriving values for the average maximum by a simulation procedure, Hintzman regards individual subject–items as having their own individual recognition and recall probability parameters. Additionally, an individual subject–item is assigned its own 2×2 contingency table parameters, so that a single subject–item can be characterized as having a certain amount of dependency between recognition and recall that can be described by a measure such as gamma. The actual observable outcome for a subject–item, a single instance of recognition success or failure combined with recall success or failure, is seen as a sample of size $n = 1$ drawn according to the hypothesized underlying and unobservable contingency table probabilities. The appropriateness of assigning contingency parameters to individual subject–items is debatable, but for the present purposes we do not question the assumption. Hintz-

Arthur J. Flexser, Department of Psychology, Florida International University; Endel Tulving, Rotman Research Institute of Baycrest Centre and University of Toronto, Toronto, Ontario, Canada.

Endel Tulving's research is supported by the Natural Sciences and Engineering Research Council of Canada, Grant A8632.

We thank Janet Metcalfe and John Wixted for constructive comments on this work.

Correspondence concerning this article should be addressed to Arthur J. Flexser, Department of Psychology, Florida International University, University Park, Miami, Florida 33199.

man's arguments, as well as our counterarguments, are substantially unchanged if we conceptualize these single-item parameters as applying to small, discrete sets of homogeneous subject–items, for which it is possible to obtain an observed 2×2 contingency table.

The average maximum, Hintzman demonstrated by simulations, depends on recognition and recall values just as the algebraic maximum does; however, unlike the algebraic maximum, the transition between cases where P(Rn) > P(Rc) and cases where P(Rn) < P(Rc) is gradual rather than sharp. This lack of a sharp transition is found because the average maximum is derived by averaging over simulated subject–items that individually show maximum dependency, with some of these subject–items characterized by recognition exceeding recall and some characterized by the reverse inequality. Because data from real experimental conditions should similarly be an average derived from a mixture of both types of outcomes occurring at the subject–item level, Hintzman argued that it is inappropriate to draw a sharp dichotomy between cases where P(Rn) > P(Rc) and cases where P(Rn) < P(Rc), as Tulving and Flexser (1992) did. Thus, as Hintzman sees it, all data points in the recognition failure scatterplot, not just those for which recall exceeds recognition, are constrained to have P(Rn|Rc) less than unity due to variability. Therefore, he argued, Tulving and Flexser's (1992) demonstration that both mathematically constrained and unconstrained experimental conditions adhere closely to the Tulving-Wiseman function is irrelevant: The unity-constrained conditions, those for which recognition exceeds recall, are actually forced below unity by virtue of the fact that many of the subject–items entering into the data for such conditions are subject to the ratio constraint.

To derive the average maximum as a function of recognition and recall, Hintzman simulated the effect of recognition and recall variability on the upper bound of P(Rn|Rc) as follows. He first selected the desired values of P(Rn) and P(Rc) for a simulated experimental condition. He then simulated individual subject–items by selecting their recognition and recall values from normal distributions with the specified means. Each simulated subject–item was assigned a gamma of unity, to assess the effect that the maximum possible dependency of each individual subject–item has on the average of all subject–items. This resulted in two distinct sets of subject–items. In both sets, the requirement that gamma equal unity resulted in one cell in the subject–item's contingency table having a zero entry. In one set, cases where recall exceeded recognition, the empty cell corresponded to recognition combined with recall failure (Rn+Rc–). In the other set, cases where recognition exceeded recall, the empty cell corresponded to recall combined with recognition failure (Rn–Rc+). A sample of size $n = 1$ was drawn according to the generated contingency table for each subject–item, and these samples were combined to produce the contingency table for the simulated experimental condition, from which P(Rn|Rc) for the condition was derived. This value of P(Rn|Rc), representing the average maximum, was considered by Hintzman to represent an upper limit, because it was based on an average of subject–items that individually were assumed to have maximum dependency. Hintzman's simulation-derived average maximum curves showed the average maximum to be appreciably lower than the algebraic maximum in many cases.

Hintzman's (1993) argument up to this point was based on the assumption that there is an appreciable amount of variability in recognition and recall across subject–items. It is this variability that figures prominently in the discussion of the concept of average maximum and its divergence from the algebraic maximum. Although the exact amount of variance that is appropriate to assume is open to debate, the presence of appreciable variability is in agreement with empirical facts and thus beyond dispute.

Serious problems, however, surface for Hintzman's (1993) conclusions when we examine a further assumption implicit in his simulations. This additional assumption is that, across subject–items, the correlation between recall and recognition is zero. The assumption is crucial, because Hintzman's whole argument is built on it: In its absence Hintzman's case collapses.

We now discuss this flawed assumption. We claim that it is contrary to fact and logically inconsistent with the objective of establishing an upper limit for P(Rn|Rc). The assumption must, therefore, be rejected. When it is replaced by a more appropriate hypothesis, the discrepancy between average maximum and algebraic maximum vanishes, and Hintzman's (1993) claim regarding the "false" dichotomy between constrained and unconstrained data itself turns out to be false.

The Crucial Assumption: Zero Across-Set Correlation

The concept of the average maximum rests on three principal assumptions. The first assumption, which we have discussed already, has to do with the existence of variability in recall and recognition across subject–items. This assumption agrees with empirical facts and is therefore acceptable.

A second assumption, also already alluded to, is that individual subject–items display the maximum possible dependency, corresponding to a gamma of unity. We refer to this dependency component as the *within-set* correlation. (This nomenclature is based on our earlier comment that it does not substantially affect the arguments to think of the individual subject–item contingency parameters as applying to small sets of homogeneous subject–items.) The justification for the assumption of a within-set correlation of unity is not empirical, but logical: To establish the upper bound of P(Rn|Rc), it is necessary to consider the case where subject–items individually show the maximum possible dependency and determine how large P(Rn|Rc) is in this case.

It is the third assumption, that recognition and recall are uncorrelated across subject–items, that we must reject. We refer to this correlation across subject–items as the *across-set* correlation. Hintzman (1993) did not explicitly acknowledge the assumption, but the lack of any statement to the contrary in his article compels us to infer that, in his simulations, he chose recognition and recall parameters for individual simulated subject–items independently. This procedure produces a situation in which the across-set correlation is zero.

The zero across-set correlation between recall and recognition that Hintzman (1993) assumed and used in his simulations is crucial. It alone among the three assumptions discussed here is responsible for creating a state of affairs in which the average maxima are lower than the algebraic maxima.

Hintzman (1993) offered no explanation for assuming zero correlation between recall and recognition across subject–items. Indeed, he paid scant attention to this third assumption. Hintzman (1993) referred to it only briefly and indirectly, and somewhat obliquely, before introducing his simulations:

> All that is needed to distort a combined contingency table is a degree of independent variation in P(RN) or P(RL) among subject–item combinations, because of subject differences, item differences, or subject–item interactions. By *independent variation*, I mean that P(RN) and P(RL) are not highly correlated. . . . In experimental terms, this means that there must be some appreciable degree of Subject × Task or Item × Task interaction. For documentation of such interactions and discussion of their effects on contingency measures, see Hintzman and Hartry (1990). (pp. 144–145)

He then proceeded to elevate the state of recall and recognition being "not highly correlated" to one of zero correlation and performed his simulations on that basis.

A fundamental problem with Hintzman's (1993) assumption of zero across-set correlation emerges in relation to the objective of establishing an upper bound on P(Rn|Rc). The upper bound represents the maximum potentially achievable value of P(Rn|Rc), because it corresponds to the situation where sources of dependency are assumed to be maximized. Consistent with this objective, Hintzman assigned a maximum gamma of unity to all simulated subject–items. That is, he assumed the within-set correlation to be maximum. However, the overall degree of dependency between recognition and recall arises from two sources: dependencies that exist within the contingency tables for the individual subject–items (within-set correlation) and dependencies that arise from recognition and recall being correlated across subject–items (across-set correlation). The assumption that the latter dependency component is zero logically guarantees that the overall dependency cannot be maximum; that is, it ensures that P(Rn|Rc) will be lower than the algebraic maximum.

There is no justification for Hintzman's treatment of the two components of dependency in exactly opposite ways: assuming the within-set correlation to be unity and assuming the across-set correlation to be zero. Available empirical evidence suggests that across-set correlation almost always occurs. Anyone who has done these kinds of experiments knows that the recognition–recall correlation between subject scores on the two tasks is substantial. For example, in the reanalysis of an experiment by Wiseman and Tulving (1975) reported by Flexser (1981), the recognition–recall correlation over subjects was found to be .71. A nonzero correlation is usually also found between item scores on the two tasks; in the same reanalysis, this correlation was found to be .28. Indeed, it was the need to correct the data of interest for such theoretically uninteresting subject and item correlations that was met by the method of "homogenization" of contingency data (Flexser, 1981).

There are also other probable sources of positive recognition–recall correlation across subject–items in addition to subject differences and item differences. Hintzman (1980) mentioned a number of sources of covariation at the subject–item level: "individual differences in vocabulary; special personal significance of items; idiosyncratic strategies; and fluctuations in attention, motivation, and mental set" (p. 400). In light of Hintzman's persistence in drawing researchers' attention to the effects such covariates can have on contingency tables (e.g., Hintzman, 1980, 1991; Hintzman & Hartry, 1990), it is odd to see him neglect them in the present context.

A Possible Rescue Attempt

Faced with this discrepancy between actual data and the assumption of zero correlation across subject–items that is incorporated into Hintzman's average maximum computations, one might be tempted to try to salvage the average maximum concept by modifying the simulation procedure to include some sort of estimated "realistic" degree of across-subject-item correlation. Conceivably, it could be argued, this neglected factor has little effect on the simulation outcomes. Furthermore, incorporating an estimated across-subject-item correlation parallels what Hintzman has already done with variance in the simulations: including what might reasonably be inferred from data to be a "realistic" amount.

However, such a rescue attempt cannot succeed. Recall that the object of the average maximum simulations is to place an upper bound on the potential recognition–recall dependency; it is not to describe actual dependency. Although estimating the across-subject-item correlation from data is clearly preferable to Hintzman's arbitrary assumption of zero correlation, an attempt to rescue the average maximum in this fashion would be flawed, because it confuses actuality and potentiality, assuming one instead of the other.

An analogy may clarify the nature of this flaw. The dependency, with its two components, might be compared with a test consisting of two parts worth 50 points each. Test takers are found typically to score about 30 points on Part 2. Would the "true" maximum possible score on this test be 80 points, on the grounds that that score is what results from an assumed perfect performance on Part 1, combined with typical observed performance on Part 2? Would one accept the argument that 100 points, instead of being the real maximum, was only a "kind of repulsive barrier"? Clearly, it is the maximum value that the second component may potentially attain, not the value that is typically obtained, that is relevant to the maximum possible total score.

In assessing the maximum dependency that can possibly occur in the recognition failure situation, it is necessary to combine Hintzman's assumption of maximum dependency within subject–items with an assumption of maximum dependency across subject–items. Only with both dependency components maximized does it make sense to compare the upper bounds of P(Rn|Rc) imposed by the average maximum and the algebraic maximum. An across-set correlation of unity need not be assumed to drive home the major conclusion of our analysis: When the proper assumptions are made, the average maximum turns out to be identical to the algebraic one. It suffices to consider the limiting case in which the upper bound of P(Rn|Rc) for all subject–items is either unity constrained, P(Rn) > P(Rc), or ratio constrained, P(Rn) < P(Rc). This contrasts with the case where subject–items represent a mixture of the two types of constraints. This limiting case is related to the correlation coefficient in that many situations with an across-subject-item correlation of unity will also produce a distribution of subject–items that are all of one constraint category or the other. For example, a recall probability for each subject–

item that is either a constant amount or a fixed proportion less than the subject–item's recognition probability yields a correlation of unity as well as a perfectly unity-constrained population of subject–items.

Proof of Equality of Average and Algebraic Maxima

We now present a simple proof that when all subject–items belong to one constraint category or the other, the average maximum coincides with the algebraic one. Recall that in deriving the average maximum, all subject–items are assumed to have gammas of unity. If all subject–items are unity constrained, then the assumption of a gamma of unity results in each subject–item's having a zero in the Rn–Rc+ contingency cell. Therefore, the contingency table for the experimental condition consisting of these subject–items will also have a zero in the Rn–Rc+ cell and will hence have a gamma of unity. A gamma of unity corresponds to the algebraic maximum, and so the condition's average maximum will coincide with the algebraic one. Similarly, if all subject–items are ratio constrained, each subject–item, and the condition as a whole, will have a zero in the Rn+Rc– contingency cell, again corresponding to a gamma of unity and an average maximum that coincides with the algebraic one.

The assumption that subject–items belong entirely to a single constraint category does not place any requirements on the values of P(Rn) and P(Rc) for the experimental condition as a whole, notwithstanding Hintzman's (1993) statement that "because of variability, it is only when P(RN) is much higher than P(RL) that the average maximum and the algebraic maximum coincide" (p. 146). Hintzman's statement is based on the flawed assumption of zero correlation across subject–items. If there is a sufficiently strong correlation between recognition and recall over subject–items, it is quite possible to have all subject–items in the same constraint category even though overall P(Rn) and P(Rc) may be arbitrarily close together and recognition and recall may be arbitrarily variable over subject–items.

Therefore, the average maximum coincides with the algebraic one for any arbitrary P(Rc) and P(Rn), provided that having all subject–items belong to the same constraint category can be considered reasonable as a potentiality. Is this outcome within the realm of possibility? We would first point out that showing that something is not impossible is a rather weak demand; the onus of proof is attached to the claim that something *is* impossible. The existence of a set of subject–items in the same constraint category is potentially as much in the realm of possibility as the existence of a set of items all of which possess a gamma of unity, the assumption embedded in Hintzman's formulation of the average maximum.

There are, in addition, theoretical reasons for considering the possibility of all subject–items belonging to the same constraint category to be a reasonable a priori assumption. Certain models predict this outcome, either generally or as a special limiting case. The simplest form of the generation–recognition model (Bahrick, 1970; Kintsch, 1970), which sees recall as having a recognition substage, predicts that P(Rn) should exceed P(Rc) for all subject–items, so that all subject–items fall into the unity-constrained category. A model that postulates that recognition and recall are based on the same trace strength characteristic, with more strength required for recall then for recognition, would make the same prediction. Both of these models are clearly falsified by empirical data, because both predict zero recognition failure. However, this discrepancy with data is irrelevant, since the average maximum is concerned with describing how dependent the data might potentially be, rather than how they actually are. If one wanted to base the argument on a model not obviously discrepant with data, however, that of Jones (1978) serves the purpose. Jones's model postulates that recall can occur through either of two processes: generation–recognition or direct access. A limiting case of this model occurs when the probability of direct-access retrieval reaches zero, when the model reduces to the generation–recognition one and recognition and recall become maximally dependent. As the probability of direct-access retrieval shrinks, an increasing proportion of subject–items fall into the unity-constrained category. Thus, the upper limit of dependency in Jones's model, corresponding to zero probability of direct-access retrieval, involves all subject–items being unity constrained, just as with the generation–recognition model.

These examples illustrate a misconception that underlies Hintzman's reasoning. In assigning recognition and recall values to simulated subject–items randomly and independently from distributions about their respective desired means, Hintzman neglected the possibility that underlying processes may place restrictions, either absolute or probabilistic, on the relative size of P(Rn) and P(Rc) for the individual subject–items. Such process restrictions may act as a source of across-subject-item correlation that exists in addition to the subject and item differences cited earlier, also neglected by Hintzman, that are reflected in correlations between subject scores and between item scores on the two tasks.

Conclusion

Hintzman argued that recognition and recall variability limit the extent of dependency between the two tasks. We agree, if such variability is uncorrelated. Inasmuch as uncorrelated variability is the antithesis of dependency, the claim that uncorrelated variability limits dependency is a tautology. This claim exactly parallels the claim that errors limit performance on a test. Hintzman's arbitrary assumption that recognition and recall have zero correlation across subject–items is equivalent to the requirement that recognition/recall dependency be lower than the theoretical maximum. The validity of Hintzman's (1993) conclusion that the average maximum imposes a stricter constraint than the algebraic one rests on the acceptance of this single crucial assumption.

We have argued that this key assumption of Hintzman's is flawed for at least three reasons: (a) It is contradicted by data that show substantial correlations between subject scores and between item scores on the two tasks, (b) it is logically inconsistent with a strategy of deriving the consequences of assumed maximization of dependency, and (c) it is at variance with a priori reasonable models.

If Hintzman's faulty assumption is replaced by a more appropriate one—that across-subject-item dependency, like within-subject-item dependency, is maximal—then, it turns out, variability does not limit dependency at all. Under these conditions, the maximum value of P(Rn|Rc) is the same as that prescribed by the algebraic maximum, corresponding to a gamma of unity. This means that the concept of the "average maximum" is redundant; it coincides with the algebraic maximum.

Given the illusory nature of the average maximum, we can restate our earlier conclusion (Tulving & Flexser, 1992): Hintzman's (1992, 1993) claim that the Tulving-Wiseman function is simply a consequence of mathematical constraints is false. The claim founders on the empirical findings that the data points adhere to the Tulving-Wiseman function regardless of the presence or absence of mathematical constraints.

References

Bahrick, H. P. (1970). Two-phase model for prompted recall. *Psychological Review, 77,* 215–222.

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

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

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

Hintzman, D. L. (1991). Contingency analyses, hypotheses, and artifacts: Reply to Flexser and to Gardiner. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 341–345.

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

Hintzman, D. L. (1993). On variability, Simpson's paradox, and the relation between recognition and recall: Reply to Tulving and Flexser. *Psychological Review, 100,* 143–148.

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.

Jones, G. V. (1978). Recognition failure and dual mechanisms in recall. *Psychological Review, 85,* 464–469.

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory* (pp. 333–374). San Diego, CA: Academic Press.

Tulving, E., & Flexser, A. J. (1992). On the nature of the Tulving-Wiseman function. *Psychological Review, 99,* 543–546.

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

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

Received August 26, 1992
Accepted September 25, 1992 ■