

List-subset effects and the Tulving-Wiseman function

CARL A. BARTLING

McNeese State University, Lake Charles, Louisiana

A current experiment, along with a reanalysis of data from an earlier study by Bartling and Thompson (1977), provided results inconsistent with the Tulving-Wiseman function (Tulving & Wiseman, 1975). With type of word pair as a within-subject variable, observed probabilities of recognition conditionalized on recall were less than predicted probabilities from the Tulving-Wiseman function for noun-adjective, cue-target pairs. The reverse was true for adjective-noun pairs. Observed probabilities of recognition conditionalized on recall were generally not different from predicted probabilities for noun-noun and adjective-adjective pairs. These list-subset effects were used to support the argument that additional research on the relation between recognition and recall is needed before supporters elevate the Tulving-Wiseman function to the status of a law.

The purpose of this study was to test whether subsets of data produced with episodic memory lists conform to the Tulving-Wiseman function (Tulving & Wiseman, 1975). Before the current research and its significance are presented, however, the phenomenon of recognition failure of recallable words and the Tulving-Wiseman function need to be described.

Under certain conditions, subjects can recall words that they have failed to recognize on a previous test of episodic memory (e.g., Bartling & Thompson, 1977; Bryant, 1991; Gardiner & Nilsson, 1990; Tulving & Thomson, 1973). This phenomenon of recognition failure of recallable words (recognition failure) is typically produced using word pairs and a sequence of tasks. This paradigm varies, but it has included (1) presentation of a list of cue-target word pairs (often low-normative associates, e.g., *ground-COLD*), (2) a free-association task in which subjects generate responses to high-normative associates of the target words (e.g., *hot*), (3) a recognition test for the target words with many of the distractors and target words having occurred as responses on the free-association task (e.g., *water, cold, boiling, burn*), and (4) a prompted-recall test for the target words with the context cues serving as retrieval prompts (e.g., *ground-?*).

Tulving and Wiseman (1975) proposed a simple quadratic function to describe the highly consistent relation between recognition and recall performance that is produced with the recognition failure paradigm (for other reviews of recognition failure studies, see Flexser & Tulving, 1978; Gardiner & Nilsson, 1990; Nilsson & Gardiner, 1991; and Nilsson, Law, & Tulving, 1988). The quadratic function described by Tulving and Wiseman is

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

where Rc refers to cued recall and Rn refers to recognition hits. Recognition failure, $P(Rn|Rc)$, is the complement of recognition conditionalized on cued recall, $P(Rn|Rc)$, in Equation 1 (cf. Tulving & Wiseman, 1975). As Tulving and his associates (e.g., Flexser & Tulving, 1978; Nilsson et al., 1988; Tulving, 1983) and others (e.g., Gardiner & Nilsson, 1990) have noted, the $P(Rn|Rc)$ observed in a wide range of experiments has consistently been very close to that predicted by the Tulving-Wiseman function.

Nilsson et al. (1988) have contended that the Tulving-Wiseman function should be elevated to the status of an empirical law because of the high degree of consistency between predictions from the function and experimental results. Conversely, exceptions to the Tulving-Wiseman function have been summarized and discussed in several articles and books (e.g., Gardiner & Nilsson, 1990; Nilsson & Gardiner, 1991; Nilsson et al., 1988; Tulving, 1983, chap. 13). One of these exceptions included the use of word pairs composed of concrete nouns and descriptive adjectives as cues and targets within the same list (Bartling & Thompson, 1977).

Nilsson et al. (1988, p. 267) emphasized that it is important to "examine the exceptions" to the Tulving-Wiseman function and "to seek their experimental and theoretical clarification." Nilsson et al. rejected many of the exceptions, including those found by Bartling and Thompson (1977), because Nilsson et al. concluded that these exceptions violated the encoding conditions specified in the "law." "The conditions under which it holds include the requirement that the to-be-remembered information T be 'encoded with respect to' the context information C" (Nilsson et al., 1988, p. 267).

Gardiner and Nilsson (1990) have contended that the results found in the Bartling and Thompson (1977) study do not deviate enough from the predicted probabilities to be considered exceptions to the Tulving-Wiseman function. This contention by Gardiner and Nilsson is based on a

Correspondence should be addressed to Carl A. Bartling, Department of Psychology, P.O. Box 91895, McNeese State University, Lake Charles, LA 70609.

Table 1
Recognition Hit Rate (Rn), Probability of Cued Recall (Rc),
Probability of Rn|Rc, and z Score for Four Types of Word Pairs
(from Bartling & Thompson, 1977, Experiment 3)

Type of Word Pair	Rn	Rc	Probability of Rn Rc			z
			Observed	Predicted	Deviation	
Noun-adjective	.61	.66	.64	.73	-.09	2.80†
Noun-noun	.70	.47	.82	.81	.01	0.58
Adjective-adjective	.76	.31	.92	.85	.07	1.84*
Adjective-noun	.79	.44	.96	.87	.09	3.03†
Average	.72	.47	.81	.82	-.01	0.61

* $p < .10$ (two-tailed test). † $p < .01$ (two-tailed test).

somewhat "arbitrary" decision as to what degree of deviation from the Tulving-Wiseman function should be considered an exception that requires further analysis and clarification. A more careful examination of the data presented by Bartling and Thompson (1977) provides another view of this exception to the Tulving-Wiseman function.

Bartling and Thompson (1977) used concrete nouns (Ns) and descriptive adjectives (As) in all four combinations as cues and targets in their study of recognition failure. The probabilities of Rn|Rc produced in their study along with predicted values from the Tulving-Wiseman function and deviations from the function for the four types of word pairs are shown in Table 1. As can be seen in Table 1, moderate deviations from the Tulving-Wiseman function were found for adjective-adjective (AA) pairs (.07), adjective-noun (AN) pairs (.09), and for noun-adjective (NA) pairs (-.09).

For 11 of the 12 NA pairs, the observed $P(Rn|Rc)$ was less than the prediction from the Tulving-Wiseman function. Conversely, for the other three types of pairs, there were 29 of 36 pairs for which the observed $P(Rn|Rc)$ exceeded the prediction from the Tulving-Wiseman function. This included 9 of 12 of the noun-noun (NN) pairs, 9 of 12 of the AA pairs, and 11 of 12 of the AN pairs. When the data are averaged across the four types of pairs, the $P(Rn|Rc)$ obtained is .81. This value is very close to the predicted value of .82, given an average recognition probability of .72 for the four types of word pairs.

Neely and Payne (1983) and Bryant (1991) have used a binomial test to evaluate the significance of the deviation observed from predicted $P(Rn|Rc)$. The binomial test uses the normal approximation to the binomial distribution with $z = (\text{observed} - \text{predicted}) / [\text{predicted}(1 - \text{predicted})/n]^{1/2}$, where observed is the observed $P(Rn|Rc)$, predicted is the predicted $P(Rn|Rc)$, and n is the number of words recalled (Bryant, 1991). This test, when used on the four types of word pairs for the Bartling and Thompson (1977, Experiment 3) study, produces the following statistics: z (NA) = -2.80, $p < .01$; z (NN) = 0.58, $p > .05$; z (AA) = 1.84, $p < .10$; and z (AN) = 3.03, $p < .01$.

An analysis of variance (ANOVA) was performed on the deviation scores [observed - predicted $P(Rn|Rc)$] for the 48 word pairs used in the Bartling and Thompson (1977, Table 3, p. 698) study. This analysis showed a sig-

nificant difference among the four types of pairs [$F(3,44) = 8.18$, $MS_e = .01$, $p < .001$]. This result lends support to the results with the binomial test reported above.

Hayman and Tulving (1989) have carried out many analyses of contingency data gathered from their laboratory over an extended period of time, but they did not find any subsets of data that would lead to different conclusions from those drawn from overall analyses. However, much of the research on recognition failure has involved homogeneous sets of word pairs of various types (e.g., NN pairs). It may be necessary to use different types of word pairs within the same list to demonstrate subset effects inconsistent with the Tulving-Wiseman function.

If it can be shown that different types of word pairs in a mixed list produce probabilities of Rn|Rc that are inconsistent with the Tulving-Wiseman function, this would be an important research finding. The Tulving-Wiseman function (law) would need to be reformulated to account for subset effects. The constant (.5) in the Tulving-Wiseman function (see Equation 1 in this paper) would need to vary for different types of word pairs.

One purpose of this study was to conduct an experiment designed to test whether subsets of data conform to the Tulving-Wiseman function. This experiment was a replication of Bartling and Thompson's (1977) Experiment 3.

Probability of Rn|Rc at the level of the experiment was expected to conform to the Tulving-Wiseman function. Probabilities of Rn|Rc for the four types of word pairs were expected to show consistent patterns of deviation from the Tulving-Wiseman function. The observed probabilities of Rn|Rc were expected to be less than the predicted probabilities for NA pairs. The reverse was expected for AN pairs. These predictions are based on results from the Bartling and Thompson (1977) article and other unpublished research by the author of the current study.

METHOD

Half of the subjects in the current study received free-choice instructions, and the other subjects received forced-choice instructions. Bartling and Thompson (1977) used instructions that were ambiguous as to the nature of their test of recognition, as they noted in their article.

Subjects

Forty-four undergraduate students in psychology courses participated in this experiment and received extra credit for their participation.

Design

The experiment was a $2 \times 2 \times 4$ mixed factorial. List (1 or 2) and type of recognition test (free choice vs. forced choice) were between-subject factors. Type of pair (NA, NN, AA, and AN) was the within-subject factor.

Materials

Most of the materials were the same as those used in Bartling and Thompson's (1977) Experiment 3. The primary change involved the use of free-choice instructions or forced-choice instructions on the recognition test for the critical list.

The two study lists (30 pairs each) were composed of six pairs of each of the four types, along with three recency and three primacy buffer pairs. The words were selected either from the Tulving and Thomson (1973) word lists or from Jenkins's 1952 Minnesota Word Association Norms (cited in Postman & Keppel, 1970) and were selected to approximate the Tulving and Thomson criteria as closely as possible. The criteria are that (a) the target word is a low-frequency (approximately 1%) associate to its list cue (e.g., *glue*-CHAIR), (b) each target word is a high-frequency (approximately 50%) associate to a corresponding word on the free-association list (e.g., *table*-CHAIR), and (c) all list words are high-frequency (of occurrence) words according to the Thorndike and Lorge (1944) norms. (Bartling & Thompson, 1977, p. 695)

Instructions, a prompted-recall task for a practice list, a free-association task, a test of recognition, and a prompted-recall task for the critical list were each presented on a single sheet of paper in a stapled test booklet with blank filler pages between task sheets. Each prompted-recall task consisted of the 24 prompts (one word from each study-list pair) for the study list. The free-association task contained the 24 high-frequency associates of the target words, along with standard free-association instructions (three responses for each high-frequency associate). The test of recognition (free choice or forced choice) consisted of 24 rows of words, with one target word and three distractors per row. The distractors were the most common free-association responses to the corresponding high-frequency associates of the target words from another experiment (Bartling & Thompson, 1977, Experiment 2).

Procedure

The subjects were tested in groups, with 4 to 6 subjects per group. The subjects were instructed to remember the words appearing together on each slide as a pair and were told that they needed to "form a distinct image (or unusual association) combining the two words in each pair, or use some other type of elaboration to help remember each of the words in a pair in relation to each other." These instructions, along with the practice list, were aimed at ensuring that subjects would encode items from each pair together. The sequence of procedures was (1) presentation of the 30-pair practice list at a 3-sec rate via slides (study cue tested in lowercase letters above the capitalized target item), (2) a 30-sec

distractor task during which subjects counted backward by threes, (3) a 3-min prompted-recall task for the practice list, (4) presentation of the 30-pair critical list at a 3-sec rate, (5) another 30-sec counting task, (6) a 6-min free-association task using the high-frequency associates of the target words, (7) a 3-min test of recognition (free choice or forced choice), and (8) a 3-min prompted-recall task for the critical list.

RESULTS

Recognition hit rate, probability of cued recall, probability of recognition conditionalized on recall, and z score by type of recognition test and type of word pair are shown in Table 2. Emphasis is on the pattern of results across conditions and not upon individual statistical tests. The reader is cautioned not to place undue emphasis on any single z score since (1) multiple z scores were used and (2) the z statistic is a relatively liberal test for these comparisons. The data are collapsed across the two lists as there were no significant main effects or any significant interactions with the list variable.

Probabilities of $R_n|R_c$ and z Scores from Group Data

Deviations from the predicted probabilities of $R_n|R_c$ resulted in the same pattern for both types of recognition tests (i.e., significant negative deviations for NA pairs, significant positive deviations for AN pairs, and nonsignificant deviations for NN and AA pairs). Results for the combined data across type of recognition test are also shown in Table 2. The observed $P(R_n|R_c)$ (.52) was less than the predicted probability (.64) by $-.12$ for the combined data for NA pairs ($z = -3.35$, $p < .01$). The reverse was true for AN pairs (.88 vs. .79; $z = 2.40$, $p < .05$). The observed $P(R_n|R_c)$ was very close (or equal) to the predicted probability for NN pairs (deviation = $-.01$) and for AA pairs (deviation = $.00$).

Deviation Scores for Individual Word Pairs

An analysis of variance was performed on the deviation scores [observed - predicted $P(R_n|R_c)$] for the 48

Table 2
Recognition Hit Rate (R_n), Probability of Cued Recall (R_c), Probability of $R_n|R_c$ (Observed and Predicted) and z Score by Condition, with Number of Responses in Parentheses

Condition	R_n	R_c	Probability of $R_n R_c$			z
			Observed	Predicted	Deviation	
Forced choice, $n = 22$.71 (375)	.48 (251)	.78 (55)	.81	-.03	-1.21
Noun-adjective	.65 (86)	.72 (95)	.65 (33)	.76	-.11	-2.66‡
Noun-noun	.70 (93)	.48 (64)	.78 (14)	.81	-.03	-0.61
Adjective-adjective	.73 (96)	.24 (32)	.87 (4)	.83	.04	0.60
Adjective-noun	.76 (100)	.45 (60)	.93 (4)	.85	.08	1.74*
Free choice, $n = 22$.52 (274)	.47 (247)	.60 (99)	.64	-.04	-1.31
Noun-adjective	.39 (51)	.64 (85)	.36 (54)	.51	-.15	-2.77‡
Noun-noun	.52 (69)	.44 (58)	.66 (20)	.64	.02	0.32
Adjective-adjective	.55 (72)	.35 (46)	.67 (15)	.67	.00	0.00
Adjective-noun	.60 (79)	.44 (58)	.83 (10)	.72	.11	1.87*
Combined, $N = 44$.61 (649)	.47 (498)	.69 (154)	.73	-.04	-2.01†
Noun-adjective	.52 (137)	.68 (180)	.52 (87)	.64	-.12	-3.35‡
Noun-noun	.61 (162)	.46 (122)	.72 (34)	.73	-.01	-0.25
Adjective-adjective	.64 (168)	.30 (78)	.76 (19)	.76	.00	0.00
Adjective-noun	.68 (179)	.45 (118)	.88 (14)	.79	.09	2.40†

* $p < .10$ (two-tailed test).

† $p < .05$ (two-tailed test).

‡ $p < .01$ (two-tailed test).

word pairs. Type of word pair approached statistical significance [$F(3,44) = 2.66$, $MS_e = 0.04$, $p < .06$]. Means and standard deviations for the four types of word pairs were: NA = $-.12$ ($SD = .08$), NN = $-.01$ ($SD = .17$), AA = $.00$ ($SD = .30$), and AN = $.09$ ($SD = .14$).

Probability of Rn|Rc from Individual Subjects

Probability of Rn|Rc was lower with the free-choice recognition test (mean = 0.63) than with the forced-choice test (mean = 0.79) [$F(1,40) = 9.19$, $MS_e = 0.11$, $p < .01$]. There was a main effect of type of word pair [$F(3,120) = 14.87$, $MS_e = 0.08$, $p < .001$]. The cell means for the four types of pairs were: NA = 0.49, NN = 0.71, AA = 0.77, and AN = 0.87. Type of recognition test did not interact with type of word pair [$F(3,120) = 0.84$, $p > .05$]. There was no main effect of list [$F(1,40) = 1.46$, $p > .05$] and no interaction between list and type of recognition test [$F(1,40) = 0.14$, $p > .05$] or between list and type of pair [$F(3,120) = 0.55$, $p > .05$].

DISCUSSION

The data from the current experiment clearly replicated the results from the Bartling and Thompson (1977, Experiment 3) study. There were subset effects for NA and AN pairs; namely, statistically significant deviations from the predicted probabilities of Rn|Rc from the Tulving-Wiseman function ($-.12$ for NA pairs and $.09$ for AN pairs). The obtained $P(Rn|Rc)$ did not significantly deviate from the predicted probability for the data collapsed across the four types of pairs for either of the two groups (free-choice vs. forced-choice test of recognition).

The degree of dependence/independence between recognition and recall is probably not a constant as Tulving and his associates (e.g., Flexser & Tulving, 1978; Gardiner & Nilsson, 1990; Nilsson et al., 1988; Tulving & Wiseman, 1975) have maintained. Instead, the degree of dependence/independence between recognition and recall probably varies as a function of other variables (e.g., the types of word pairs used with the recognition failure paradigm). Hintzman (1980) had earlier advocated this general position, but his focus was on demonstrating the importance of Simpson's paradox with contingency analyses through the use of hypothetical data. Bartling and Thompson (1977) presented empirical data that could be used to suggest possible subset effects for different types of word pairs. However, their focus had been on predicting the probability of recognition failure with an empirically derived measure of retrieval asymmetry (cf. Salzberg, 1976, for more information on the relation between retrieval asymmetry and recognition failure).

Hintzman and his colleagues (e.g., Hintzman & Hartry, 1990) have argued that the Tulving-Wiseman function is an artifact, with its regularity due to a combination of mathematical constraints, typical data values, and suppressor variables. Their view is sharply contrasted with the view of Tulving and others (e.g., Gardiner & Nilsson, 1990; Nilsson & Gardiner, 1991; Nilsson et al., 1988) who have advocated raising the status of the Tulving-Wiseman function to a law.

The current study does not attempt to resolve this conflict of opposing views. Obviously, there is some degree of dependence between recognition and recall performance. This dependence is probably neither a constant across all conditions nor simply an artifact. However, the mathematical constraints and related factors discussed by Hintzman and Hartry (1990) may make it difficult to demonstrate empirically that variables such as type of word pair (e.g., noun-adjective pairs vs. adjective-noun pairs) influence the degree of dependence between recognition and recall performance.

The author of the current study can offer only conjecture as to why different types of word pairs differ in their degree of dependence between recognition and recall. Different types of word pairs may differ in the extent that automatic versus controlled retrieval processes are generated to access the episodic memory context to permit recognition or recall of a target word. This idea is consistent with subjects' self-reports concerning the ease of recalling target words when provided the corresponding context cues. For example, subjects report that it is often "effortless" to recall a descriptive adjective (e.g., *COLD*) when provided the concrete noun (e.g., *ground*) that served as the original encoding context. The reverse is true when subjects are provided a descriptive adjective and asked to recognize it or use it as a retrieval cue to assist in the recall of a concrete noun that was presented with it.

REFERENCES

- BARTLING, C. A., & THOMPSON, C. P. (1977). Encoding specificity: Retrieval asymmetry in the recognition failure paradigm. *Journal of Experimental Psychology: Human Learning & Memory*, 3, 690-700.
- BRYANT, D. J. (1991). Exceptions to recognition failure as a function of the encoded association between cue and target. *Memory & Cognition*, 19, 210-219.
- FLEXSER, A. J., & TULVING, E. (1978). Retrieval independence in recognition and recall. *Psychological Review*, 85, 153-171.
- GARDINER, J. M., & NILSSON, L.-G. (1990). Relation between recognition and recall: The Tulving-Wiseman law. Manuscript submitted for publication.
- HAYMAN, C. A. G., & TULVING, E. (1989). Contingent dissociation between recognition and fragment completion: The method of triangulation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15, 228-240.
- 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 and fragment completion: Contingency relations vary for different subsets of words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16, 955-969.
- NEELY, J. H., & PAYNE, D. G. (1983). A direct comparison of recognition failure rates for recallable names in episodic and semantic memory tests. *Memory & Cognition*, 11, 161-171.
- NILSSON, L.-G., & 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* (pp. 57-74). Hillsdale, NJ: Erlbaum.
- 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, & Cognition*, 14, 266-277.
- POSTMAN, L., & KEPPEL, G. (1970). *Norms of word association*. New York: Academic Press.
- SALZBERG, P. M. (1976). On the generality of encoding specificity. *Journal of Experimental Psychology: Human Learning & Memory*, 2, 586-596.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University Teachers College Bureau of Publications.
- TULVING, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- TULVING, E., & THOMSON, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373.
- TULVING, E., & WISEMAN, S. (1975). Relation between recognition and recognition failure of recallable words. *Bulletin of the Psychonomic Society*, 6, 79-82.