JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 21, 237–248 (1982)

Priming and Recognition Failure

ARTHUR J. FLEXSER

Florida International University

AND

ENDEL TULVING

University of Toronto

A theoretical account of the priming effects found in the recognition-failure paradigm is proposed. The recently established existence of priming effects in this paradigm posed a problem for the Flexser and Tulving (*Psychological Review*, 1978, 85, 153–171) model, which ignored priming. We present a revised version of the model that incorporates priming; it does so at the level of individual constituent features of the episodic memory trace rather than at the level of whole words. Among other things, such a view implies that unrecognized as well as recognized items are subject to priming. When priming is taken into account in this fashion, the small discrepancy that existed between our original model and the data is further reduced.

When words presented as members of pairs are subsequently tested successively for recognition and for cued recall, instances commonly occur in which nonrecognizable words are successfully recalled by the same subject. This phenomenon of *recognition failure of recallable words* (or simply *recognition failure*) has been the subject of a number of empirical and theoretical investigations (e.g., Begg, 1979; Jones, 1978; Kintsch, 1978; Rabinowitz, Mandler, & Barsalou, 1977; Vining & Nelson, 1979; Watkins & Tulving, 1975). One proposed explanation of the existence of recognition failure traces the phenomenon to the independent operation of different retrieval cues in accessing the episodic memory trace corresponding to the presentation of a word pair (Tulving & Thomson, 1973; Watkins & Tulving, 1975; Wiseman & Tulving, 1976). In recognition, a copy of the target item acts as a retrieval cue, while in recall the retrieval cue corresponds to the other member of the original word pair. In both cases, however, the same episodic memory trace is assumed to be accessed.

This account was formalized with some added assumptions by Flexser and Tulving (1978) and specified in the form of a mathematical model. The model was proposed to account for an empirical relation pointed out by Tulving and Wiseman (1975) between recognition probability, $p(\text{Rn})$, and the probability of recognition conditionalized on successful recall, $p(\text{Rn}|\text{Rc})$. (The latter quantity is the complement of the recognition-failure probability.) Tulving and Wiseman (1975) presented a scatterplot showing $p(\text{Rn}|\text{Rc})$ plotted against $p(\text{Rn})$ for several dozen experimental conditions, representing work published by a number of investigators. On such a scatterplot, the positive diagonal represents independence and the upper edge represents complete dependence. All of the plotted conditions conformed to the basic recognition-failure paradigm in which words presented as pairs

Requests for reprints should be sent to Arthur J. Flexser, Department of Psychology, Florida International University, Tamiami Campus, Miami, Fla, 33199. Portions of this research were supported by National Research Council of Canada Grant A8632 to Endel Tulving.

0022-5371/82/030237-12$02.00/0

were tested successively for recognition and cued recall. Beyond this common aspect, however, there was wide procedural diversity with regard to materials, instructions, and other variables.

Notwithstanding this diversity, the data points in Tulving and Wiseman's (1975) scatterplot proved to be surprisingly concentrated in a relatively narrow bowed-shaped region lying a moderate distance above the independence diagonal, representing a moderate degree of dependency between recognition and recall outcomes. The equation $p(\text{Rn}|\text{Rc}) = p(\text{Rn}) + c[p(\text{Rn}) - p(\text{Rn})^2]$, with $c = .5$, was found by Tulving and Wiseman to provide a good empirical description of the region of maximum density on the scatterplot.

Flexser and Tulving (1978) proposed a mathematical model to account qualitatively and quantitatively for the pattern of results pointed out by Tulving and Wiseman (1975). The encoded versions of the memory trace and the retrieval cues were represented as collections of features, with successful retrieval being assumed to occur when the number of feature matches between the retrieval cue and the memory trace exceeded a criterion level. Crucial to the model was the assumption of *retrieval independence*, which stated that the features encoded for the retrieval cues on the two retrieval occasions were statistically independent. That is, whether or not a feature was encoded for the recognition cue had no effect on the corresponding feature's probability of being encoded for the recall cue. This assumption would lead to independence between recognition and recall outcomes were it not for the additional assumption of *variability in goodness of encoding*. Dependency between outcomes was assumed to arise because the memory traces for some study pairs were better encoded (i.e., had more encoded features) than for others. Since the same trace was assumed to be accessed by both retrieval cues, variability in goodness of encoding would result in correlated outcomes on the recognition and recall tests.

Flexser and Tulving's (1978) model employed six parameters: N, the dimensionality of the feature space (i.e., the number of features necessary to specify any possible memory trace, seen as a fixed property of the memory system); p, the probability that any given feature is successfully encoded as a part of the episodic trace; r and s, the encoding probabilities for features in the recognition and recall cues, respectively; and the criterion parameters I and J, representing the number of cue-to-trace feature matches necessary for a successful response to occur in recognition and recall, respectively.

Based on this framework, Flexser and Tulving (1978) derived expressions for $p(\text{Rn})$, $p(\text{Rc})$, and $p(\text{Rn}|\text{Rc})$ in terms of the model parameters. The diverse experimental conditions for which data were available were assumed to be representable by diverse parameter combinations. Thus, an arbitrarily selected set of parameters could be thought of as simulating some (unknown) set of experimental circumstances within the recognition-failure paradigm for which a simulated data point on the recognition-failure scatterplot could be generated using the model equations. Since the recognition-failure scatterplot from real experiments represented a pattern of outcomes from diverse experimental circumstances, a test of the model would be whether it yielded the same pattern of outcomes on a simulated recognition-failure scatterplot consisting of data points generated by selecting sets of parameter values randomly, subject to certain reasonable constraints.

Figure 1, taken from Flexser and Tulving (1978), illustrates the outcome of such a test of the model, showing the comparison between real data (top panels) and model-generated data (bottom panels) for recognition-failure scatterplots (left) and for scatterplots of overall recognition and recall

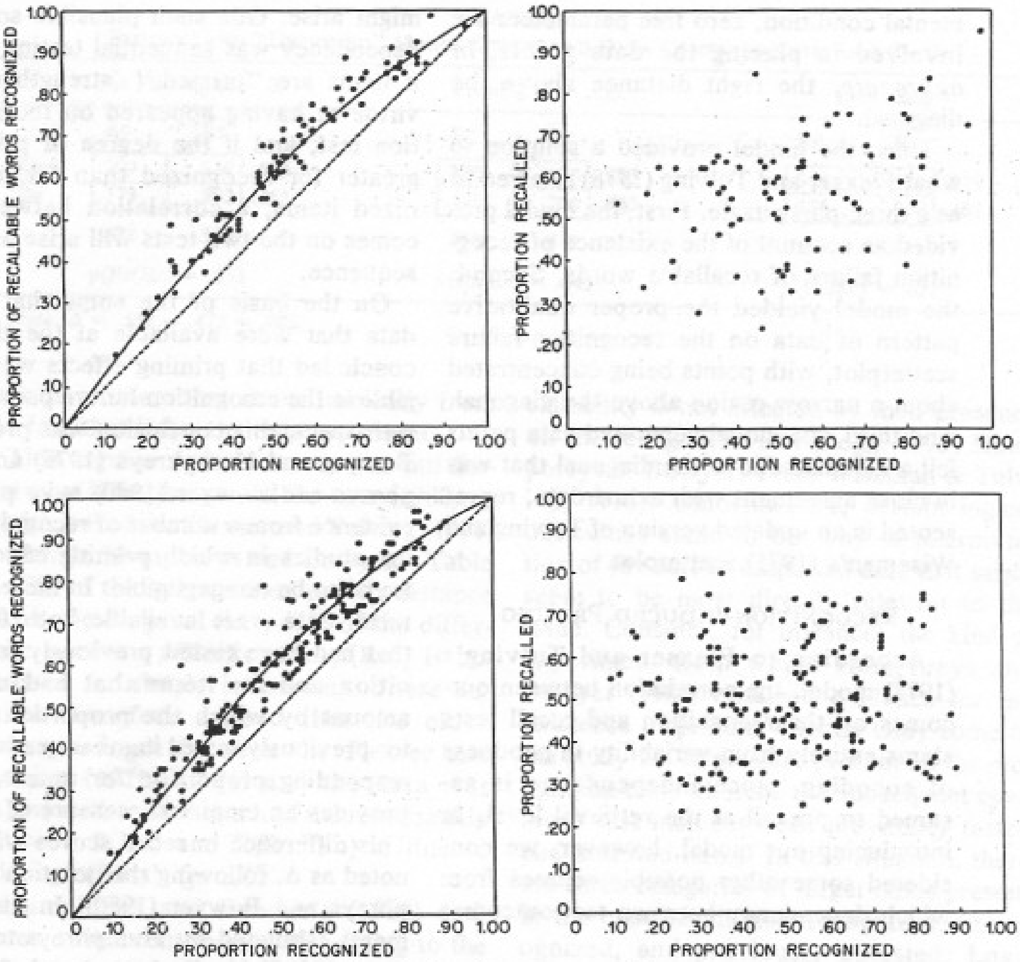

FIG. 1. Four scatterplots depicting relations between recognition and recall. The upper two graphs represent real data from 89 conditions in 33 experiments; the lower two represent model-generated data from 200 simulated experimental conditions. The graphs on the left depict recognition failure; those on the right show the relation between overall recall and recognition scores. (From Flexser & Tulving, 1978.)

levels (right). Visual inspection suffices to confirm that the model performs very creditably in capturing the essential features of the recognition-failure scatterplot under conditions where recognition and recall levels vary widely and in largely uncorrelated fashion.

It should be emphasized that the quantitative fit of the model, in terms of placing the data points the proper distance above the diagonal, was a consequence of the model's structure and did not depend on any fixed parameter input. The points simply fell in the proper region when all parameters were allowed to vary randomly.[1] In other words, although the model requires six parameters to describe a single experi-

[1] *N* was kept fixed in generating simulated data points, reflecting this parameter's role in corresponding to a fixed property of the memory system. Because the role of *N* diminishes strongly as *N* increases, the value assigned to this parameter is largely irrelevant so long as *N* exceeds 10. The value of 20 was used for the simulation to be described later.

mental condition, zero free parameters are involved in placing the data points, *in aggregate*, the right distance above the diagonal.[2]

Thus, the model provided a solution to what Flexser and Tulving (1978) referred to as a three-part puzzle. First, the model provided an account of the existence of recognition failure of recallable words. Second, the model yielded the proper qualitative pattern of data on the recognition-failure scatterplot, with points being concentrated about a narrow region above the diagonal. And third, the model-generated data points fell a distance above the diagonal that was in close agreement with actual data, represented in an updated version of Tulving and Wiseman's (1975) scatterplot.

Recognition-Induced Priming

According to Flexser and Tulving's (1978) model, the correlation between outcomes on the recognition and recall tests stems entirely from variability in goodness of encoding, since independence is assumed to prevail at the retrieval level. In introducing our model, however, we considered some other possible sources from which dependency between test outcomes might arise. One such plausible source of dependency was sequential testing effects: if items are "primed," (strengthened) by virtue of having appeared on the recognition test, and if the degree of priming is greater for recognized than for unrecognized items, a correlation between outcomes on the two tests will arise as a consequence.

On the basis of the somewhat meager data that were available at the time, we concluded that priming effects were negligible in the recognition-failure paradigm. In retrospect, this conclusion was premature. Bowyer and Humphreys (1979) and Humphreys and Bowyer (1980) have presented evidence from a number of recognition-failure studies in which priming effects were found to be nonnegligible. In these studies, recall data were available both for items that had been tested previously for recognition and for items that had not. The amount by which the proportion recalled for previously tested items exceeds the corresponding proportion for untested items provides an empirical measure of priming. This difference in recall scores will be denoted as Δ, following the notation of Humphreys and Bowyer (1980). In the experiments tabulated by Humphreys and Bowyer (1980, Table 1), Δ averaged .08, with 9 out of 16 experimental conditions cited showing a significant priming effect.

These findings make it plausible that sequential testing effects account for some of the dependency between recognition and recall outcomes that is observed in the recognition-failure paradigm. They also create a problem for the Flexser and Tulving (1978) model: If priming acts as an additional source of dependency not represented in the model, the model in neglecting this factor should have appreciably underpredicted the amount of dependency present in real data, and the model-generated scatterplot should seemingly have fallen appreciably closer to the diagonal than did the real-data scatterplot. Thus, the Flexser and Tulving (1978) model seems to have a

[2] Hintzman (1980) has suggested that the distance of the simulated data points from the diagonal depends on a "hidden" parameter, namely, the range over which p was varied in generating scatterplots, which in this case was .20 to .80. Hintzman's argument is erroneous since, for each data point, the level of recognition failure depends only on the value of p applying to that particular data point and is independent of the range over which p varied for other data points. Hintzman apparently has confused the parameter p, which is fixed for a particular experimental condition, with n, the number of features encoded for an individual simulated subject-item within an experimental condition. The within-experimental variability in n, which is binomially distributed with variance $Np(1 - p)$, is an important determinant of recognition-failure levels, but the range over which p is varied in simulating experiments is not. For example, when the p range was narrowed from .05–.95 to .45–.55, the average deviation of the data points from the Tulving and Wiseman (1975) function changed only slightly, from −.006 to −.004.

TABLE 1

"PRIMED" AND "UNPRIMED" MODEL-GENERATED DATA COMPARED WITH REAL DATA

	No Priming (k = 0)			Priming (k = .17)			Real Data		
	Rc	$\overline{\text{Rc}}$		Rc	$\overline{\text{Rc}}$		Rc	$\overline{\text{Rc}}$	
Rn	.299	.236	.535	.359	.176	.535	.364	.171	.535
$\overline{\text{Rn}}$	.173	.292	.465	.199	.266	.465	.195	.270	.465
	.472	.528		.558	.442		.559	.441	
	p(Rn|Rc) = .633			p(Rn|Rc) = .643			p(Rn|Rc) = .651		

Note. Δ = .558 − .472 = .086.

serious problem that stems ironically from the fact that it fits the data too closely.

This argument, in essence, was put forward by Humphreys and Bowyer (1980) as a criticism of the Flexser and Tulving (1978) model. In Humphreys and Bowyer's Table 1, 47% of the average data point's distance above the diagonal [i.e., 47% of the difference between p(Rn|Rc) and p(Rn)] is accounted for simply by priming. These authors therefore argued that the close correspondence between the Flexser and Tulving (1978) model and the data was "spurious because much of the relationship between P(Rn) and P(Rn | Rc) is due to priming effects" (p. 257).

Humphreys and Bowyer's (1980) estimates of the contribution of priming to the observed dependency were, however, based on the assumption that priming involves only those words that are recognized in the test. As we shall show, this assumption is highly critical. We will examine the evidence relevant to the assumption next, and conclude that the assumption is not tenable: Priming occurs for both recognized and unrecognized words. We will then present a revised version of our 1978 model that incorporates priming effects.

EVIDENCE CONCERNING PRIMING OF UNRECOGNIZED WORDS

To this date, all writers who have been concerned with the problem of the role of priming effects in recognition failure have assumed that unrecognized items are not primed, that is, that their subsequent recallability is not affected by their presence on the recognition test (Begg, 1979; Humphreys & Bowyer, 1980; Wiseman & Tulving, 1976). Unfortunately, certain logical difficulties stand in the way of interpretation of the sort of empirical data that might seem to be most directly relevant to the issue. Consider, for instance, the kind of experiment from which Humphreys and Bowyer (1980) obtained evidence for the existence of priming effects: Only some of the study-list targets are present on the recognition test, whereas the subsequent cued recall test includes both previously tested and untested items. In this situation, there are three categories of target items present in the cued-recall test: recognized, unrecognized, and previously untested. Logically, the assumption that unrecognized items are not primed implies that the probability of recall of unrecognized items is identical with that of untested items. Virtually all published data, however, clearly show that the probability of recall of untested items is higher than the probability of recall of unrecognized items (e.g., Begg, 1979, Experiment 1; Bowyer & Humphreys, 1979; Postman, 1975).[3]

This sort of outcome can be interpreted as reflecting item selection effects, in the

[3] Several conditions of Experiment 2 of Begg (1979) form an exception to this general finding. However, the theoretical significance of this exception is in doubt due to the possibility, pointed out by Humphreys and Bowyer (1980, p. 276), that the finding is an artifact of the within-subject instructional manipulation employed by Begg.

same way in which the higher recall of recognized items than of untested items could be thought to reflect the combined effects of item selection and priming. The fact that recall of unrecognized items is lower than recall of untested items does not rule out the possibility that unrecognized items are primed: The lower recall of unrecognized than of untested items may simply mean that the positive priming effects are outweighed by the negative item selection effects. But since no methods have been proposed for separating the two kinds of effects, the previously available data are incapable of determining whether or not unrecognized items are subject to priming.

Data relevant to the issue, however, have been found in two recent experiments done at Toronto with the collaboration of Ruth Donnelly. In these experiments, which conformed to the standard recognition-failure paradigm, subjects were tested after three different retention intervals that included longer periods of time than have been used in most previous experiments. In the second of the two experiments, for instance, subjects were tested for both recognition and cued recall immediately, at 7 days, and at 28 days. The materials used were phrase–word pairs such as MATES LIVE IN, EMPLOYEES OUTSIDE—PRISON, which have been found to yield relatively high retention at long delays (Tulving & Watkins, 1977). The total set of study list words was divided into three nonoverlapping subsets, each of which constituted the target set on a particular retrieval occasion. At each retention interval, only one-half of the target words from the relevant subset appeared on the recognition test whereas cues for all the target words in the relevant subset were presented in the immediately following cued recall test. Thus, it was possible to compare cued recall (scored stringently) for recognized, unrecognized, and untested target items after different retention intervals.

The data from the immediate tests clearly conform to the pattern that has been reported by others: Recall was highest for the recognized words, intermediate for untested words, and lowest for unrecognized words. As the retention interval increased, however, the relation between the untested and unrecognized words reversed. After the longest retention interval in each of the experiments—14 days in the first and 28 days in the second experiment—recall of unrecognized words was clearly higher than the recall of untested words. Thus, for instance, after the 28-day interval in Experiment 2, probability of recall for recognized items was .43, for unrecognized items .27, and for untested items .16.

This kind of outcome—higher recall of unrecognized than of untested target words—provides direct support for the assumption that unrecognized items may be primed. Since item selection effects would produce a difference in the opposite direction, the observed difference can reasonably be attributed to the priming of unrecognized items. Item selection effects in this case can be expected simply to attenuate the observed priming effects to some degree.

In addition to the empirical evidence just presented, there are also theoretical reasons for supposing that unrecognized items are subject to priming, contrary to the assumption of Wiseman and Tulving (1976) and others. There is at present a great deal of evidence, based on signal-detection analysis and other considerations, that suggests that recognition is most appropriately viewed as a continuous rather than an all-or-none process. Although the subject may be asked to make discrete categorization decisions in the recognition test, there are strong reasons for thinking of the underlying informational base for these decisions as a continuum. Within this viewpoint, it simply does not make good sense to argue that priming occurs for test items falling on one side of the decision criterion in the underlying distribution of relevant evidence and not for items falling on the other side.

Rather, it seems more reasonable to suppose that the degree to which priming occurs should be largely independent of the subject's decision biases and overt recognition responses, and should instead depend on processes more closely allied to the operation of the memory system per se. In terms of the General Abstract Processing System described by Tulving (in press), priming can be regarded as a form of recoding of the encoded engram of an event; it is not dependent upon the process of conversion of ecphoric information into an overt response.[4] Among other things, thinking about priming in this way allows for the accommodation of results from the confidence judgment recognition procedure, in which there is no sharp distinction between recognized and unrecognized items.

Priming of Trace Features

Since priming induced by the recognition test is an established fact, and since the original Flexser and Tulving (1978) model made no provisions for it, the model in its original form must be regarded as inadequate, and a revision that takes priming into account is necessary. One way of accommodating the mechanism of priming is to assume that priming does occur and that it does so at the level of individual features that are useful for recognition, rather than at the level of the target word itself. In the model, encoding a feature is equivalent to assigning it a value. In order to be useful for retrieval, a feature must have been encoded with matching values on the study and retrieval occasions; successful retrieval occurs when the number of such useful features equals or exceeds some criterion number. Thus, in recognition, a hit is assumed to occur when the number of useful features (i.e., those matching between the copy cue and the encoded trace) equals or exceeds a criterion value.

The specific assumption that we now wish to add to the model with respect to priming is that a feature of the encoded trace that is categorized as useful for recognition has an enhanced probability of being useful for recall. There are a number of possible mechanisms by which this might be accomplished, and the specification of the exact nature of such a mechanism is unnecessary for purposes of deriving predictions from the model. One possibility, offered in a speculative spirit, relies on the resonance metaphor discussed in our earlier paper (Flexser & Tulving, 1978, p. 170). Within this metaphor, when an item is tested for recognition, matching features between cue and trace resonate with one another by virtue of their matched values, analogous to the phenomenon of a struck tuning fork producing resonance in a second fork tuned to the same frequency. Recognition testing of an item might be seen as leaving behind temporary residual resonance, or activation, in those features of the trace that were useful for recognition (i.e., those features that matched corresponding features in the encoded copy cue). According to the model, corresponding feature values must match between the encoded trace and the retrieval cue to within some tolerance in order for a feature to be useful. Suppose that the effect of prior activation of a feature were to widen that feature's tolerance interval for a match, that is, to "sensitize" it. Such a primed feature would then have an enhanced probability of being useful for recall.

It should be noted explicitly that the revised model assumes that priming affects all useful (i.e., matched) recognition features, regardless of whether the number of such

[4] Since we know very little at present about the properties of priming and the processes underlying it, we should not entirely rule out the possibility that the actual emitting of a particular response in the recognition test, whether positive or negative, is somehow important, perhaps by way of influencing the extent and nature of additional processing accorded to the item. For our present purposes, however, we choose to assume that the extent of priming depends on the number of matching features between trace and cue, and that it does not depend crucially on processes associated with the overt production of the response.

features exceeds the recognition criterion—that is, whether or not recognition actually occurred. Thus, an unsuccessful recognition attempt is assumed to influence later recall probability by exactly the same process by which a successful attempt does so—that is, by the activation of matched features in the episodic memory trace.

With the added assumption that priming has the effect of giving features that were useful for recognition an enhanced probability of being useful for recall, the revised model conveniently turns out to be identical to the "general" model discussed by Flexser and Tulving (1978). This "general" model was introduced to demonstrate the consequences of relaxing the retrieval independence assumption embodied in our principal model, which was also referred to as the "special" model. In the general model, the single parameter s, which in the special model represented the encoding probability for features in the recall cue, was replaced by the two parameters s_1 and s_2, representing encoding probabilities for recall cue features that were respectively useful or not useful in recognition. In applying this model to priming, we will let s_1 represent the recall encoding probability for features primed by the recognition retrieval attempt (regardless of whether or not a recognition hit actually occurred), and let a lesser value of s_2 represent the encoding probability for unprimed features of the recall cue.

To investigate systematically the effects of introducing varying degrees of dependency, Flexser and Tulving (1978) generated scatterplots in which s_2 was constrained to be a fraction of s_1, with the latter quantity varying randomly over parameter combinations. This relation was specified in the equation $s_2 = (1 - k)s_1$, where k was fixed for a given scatterplot. When k equals zero, $s_1 = s_2$ and the general model reduces to the special one embodying the assumption of retrieval independence. Greater values of k represent greater effects of a feature having been previously useful for recognition on the probability of that feature's being useful for recall, and are accompanied by a greater degree of dependency between recognition and recall outcomes, seen on the recognition-failure scatterplot as a raising of the data points in relation to the independence diagonal. The best-fitting value of k was quite small, causing Flexser and Tulving (1978) to conclude that the special model, embodying retrieval independence, represented an appropriate description of the recognition-failure paradigm.

Based on subsequent considerations, it now seems that retrieval independence applies strictly only in situations where priming does not occur, with the general model applicable in other instances. With these assumptions, the special and general models can be used in conjunction with one another to investigate the effects of priming within individual simulated experimental conditions. That is, we can compare two versions of the outcome of a particular simulated experimental condition that corresponds to a set of parameters in the general model. One version represents a simulation of a situation in which priming exists, equivalent to choosing a set of parameter values in the general model for which the relation $s_1 > s_2$ applies. The other version represents a simulation of a hypothetical situation identical with the previous one except that priming effects are nonexistent. In this model, this is equivalent to choosing the identical set of parameter values as before with the exception of s_1, which is now set equal to s_2. This latter case is one in which the general model reduces to the special one, where retrieval independence applies.

Testing of the Revised Model

As noted above, the revised model allows the generation of two versions of the 2×2 table of outcome contingencies pertaining to a particular simulated experimental condition. By computing the difference between recall levels in the "primed" and "unprimed" cases, we obtain a simulated equivalent of Δ, the observable measure of

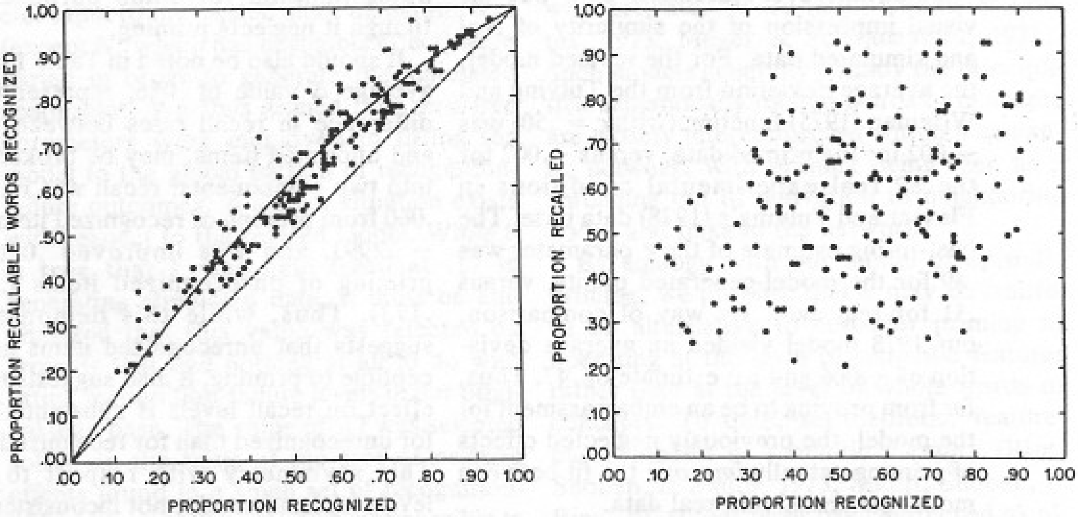

FIG. 2. Two scatterplots depicting relations between recognition and recall, generated by the revised model in 200 simulated experiments. The graph on the left depicts recognition failure; that on the right shows the relation between overall recall and recognition scores.

the degree of priming in an experimental condition. To test the revised model, it was necessary to generate a sample of simulated data points for which the average Δ approximated that found in real experiments, thus equating the average degree of priming present in the real and simulated cases.

In order to match the experimentally observed degree of priming, we adopted the technique used by Flexser and Tulving (1978) of generating scatterplots corresponding to different values of k. The value of k that applied to a particular scatterplot represented an index of the average magnitude of the priming effect; when k equals zero, there are no priming effects and retrieval independence prevails. For present purposes, the relevant consideration was that, by choosing k appropriately, we could generate scatterplots characterized by any desired average Δ value. We therefore determined by trial and error that a value of $k = .17$ produced an average Δ value of .086, in reasonable agreement with the corresponding value of .080 that applies to the real experiments cited in Humphreys and Bowyer's (1980) Table 1. This allowed the critical test to be made: Would the model, with $k = .17$, still place the data points the proper distance above the diagonal on the recognition-failure scatterplot?

The test of the model in its revised form involved 1000 simulated experimental conditions in which primed ($k = .17$) and unprimed ($k = 0$) versions of the 2 × 2 contingency tables of outcomes were generated.[5] Figure 2 shows the revised model's primed versions of the recognition-failure scatterplot and the plot of p(Rc) versus p(Rn). Both are in striking agreement with the corresponding plots for real data shown in the upper panels of Figure 1. (Only the first 200 of the 1000 simulated experimental conditions are shown in Figure 2 in order to make these plots visually more easily comparable to the earlier ones.)

[5] For purposes of matching certain global properties of the real and simulated data more closely, our original selection procedures for the criterion parameters I and J were modified somewhat. These modifications provided a match to within .01 for mean recognition and recall levels (.54 and .56 respectively), and for the across-experiment correlation between overall recognition and recall scores ($r = .29$). As noted by Flexser and Tulving (1978, p. 169), this correlation should not be confused with the crucial within-experiment correlation (phi coefficient) that determines a data point's location with respect to the independence diagonal. Details of the modified parameter selection procedures are available from the first author on request.

Quantitative comparisons bear out the visual impression of the similarity of real and simulated data. For the revised model, the average deviation from the Tulving and Wiseman (1975) function (with $c = .50$) was $-.002$ for the primed data, versus $+.002$ for the 89 real experimental conditions in Flexser and Tulving's (1978) data base. The best-fitting estimate of the c parameter was .49 for the model-generated points, versus .51 for real data. By way of comparison, our 1978 model yielded an average deviation of $-.006$ and a c estimate of .47. Thus, far from proving to be an embarrassment for the model, the previously neglected effects of priming actually *improve* the fit between model-generated and real data.

Table 1 provides an illustration of the effects of priming as simulated by the model. This table illustrates the "typical" outcome, obtained by averaging together, cell by cell, the contingency tables from our 1000 simulated experiments for the first two panels, and from the 89 real experimental conditions for the third panel.

Note first how small is the change in $p(\text{Rn}|\text{Rc})$ produced by priming—that is, how little the location of the data point corresponding to this "typical" simulated experiment is altered by the presence of an appreciable amount of priming. This is contrary to expectations that one would form based on Humphreys and Bowyer's (1980) arguments, and is largely a consequence of introducing the assumption that unrecognized items are subject to priming: While priming of recognized items raises the points on the recognition-failure scatterplot (i.e., decreases the amount of recognition failure), priming of unrecognized items increases the recognition failure rate, lowering the data points. As the "typical" outcome in Table 1 shows, these two processes largely offset each other, resulting in little net effect of priming on recognition failure rates. This insensitivity of recognition-failure levels to priming implies that our 1978 model, embodying retrieval independence, may be suitable for use as an approximation for some purposes even though it neglects priming.

It should also be noted in Table 1 that the average Δ value of .086, representing the difference in recall rates between primed and unprimed items, may be broken down into two components: recall was improved .060 from priming of recognized items (.359 − .299), and was improved .026 from priming of unrecognized items (.199 − .173). Thus, while this demonstration suggests that unrecognized items are susceptible to priming, it also suggests that the effect on recall levels is substantially less for unrecognized than for recognized items. This asymmetry with respect to *recall* levels is, incidentally, not inconsistent with the fact that priming effects on recognized and unrecognized items roughly cancel each other with respect to *recognition failure* levels, as Table 1 shows.

Several aspects of the simulated data that have not yet been mentioned are also in good agreement with real data. First, the correlation between Δ and deviations from the Tulving and Wiseman (1975) function was negligible both for real data (Table 1 in Humphreys and Bowyer, 1980), where $r = .02$, and for model-generated data, where $r = -.04$. This finding is consistent with the insensitivity of recognition-failure levels to priming. Second, Humphreys and Bowyer (1980) pointed out that a positive correlation of .86 existed between Δ and $p(\text{Rn})$ in the experiments they cited. A positive correlation also exists in the model-generated data, where, for the 1000 simulated conditions, $r = .58$. Finally, the ranges of Δ values for real data (−.08 to .19) and for simulated data (.02 to .24) are in reasonable agreement, especially if two cases of nonsignificant negative priming effects in the real data (−.08 and −.01) are regarded as instances of measurement error.

When we earlier discussed the fit of the original Flexser and Tulving (1978) model to the recognition-failure scatterplot we noted that the close agreement obtained was not a consequence of any fixed parameter input to the model, since all parameters (except N, which has negligible effect) were varied in order to generate the simulated scatterplot. We thus argued that zero free parameters were employed in fitting the model to the global pattern of recognition-failure outcomes. A similar situation exists with respect to the revised model. While it is true that k was held fixed at .17 in generating simulated data, it must be emphasized that this value was selected on grounds that had nothing at all to do with fitting recognition-failure levels in our original data base. The value of .17 was selected solely on the basis of the size of priming effects found in a small set of recognition-failure experiments where it was possible to estimate priming.

Conclusion

It was stated that the original Flexser and Tulving (1978) model solved a three-part puzzle in accounting for (a) the existence of the phenomenon of recognition failure, (b) the qualitative constancy in the recognition-failure scatterplot, and (c) the quantitative location of the data points with respect to the diagonal. The new model presented here suggests a solution to a five-part puzzle, in that the model now also accounts for (d) the existence of recognition-induced priming effects, and (e) the fact that experimental conditions showing both large and small priming effects yield data points in the same bowed-shaped region of the recognition-failure graph, with negligible dependence of recognition-failure levels on priming levels.

Finally, we wish to note the possible applicability of the type of priming mechanism we have discussed to other types of priming situations, specifically those involving semantic, rather than episodic priming. A number of investigators have proposed "spreading activation" models of semantic priming to account for facilitation effects observed in various tasks involving the testing of words bearing a semantic relationship to a previously tested (or otherwise activated) word (e.g., Collins & Loftus, 1975; Meyer & Schvaneveldt, 1971). Such models generally employ the principle that semantically related words or concepts are closely interconnected within a semantic network, with priming being seen as a spreading out of activation to neighboring nodes and links in the network.

By analogy with our model of episodic priming, we propose that it may be fruitful as an alternative to consider priming as operating at the level of semantic features rather than at the level of whole words or concepts. To date, set-theoretic, feature-based models of semantic memory (Rips, Shoben, & Smith, 1973; Smith, Shoben, & Rips, 1974) that have been proposed as alternatives to network models have had little to say about priming. Within the framework of such models, we suggest that priming could be viewed as occurring through activation of features that are shared between related concepts. While such a viewpoint can readily be translated into a network representation (see Hollan, 1975), it may well develop that the feature representation of priming can suggest useful alternative processing assumptions that would be unlikely to emerge within the network viewpoint (see Rips, Smith, & Shoben, 1975).

References

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

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

Collins, A. M., & Loftus, E. F. A spreading activation theory of semantic processing. *Psychological Review*, 1975, 82, 407–428.

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

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

Hollan, J. D. Features and semantic memory: Set-theoretic or network model? *Psychological Review*, 1975, 82, 154–155.

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

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

KINTSCH, W. More on recognition failure of recallable words: Implications for generation-recognition models. *Psychological Review,* 1978, 85, 470–473.

MEYER, D. E., & SCHVANEVELDT, R. W. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology,* 1971, 90, 227–234.

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

RABINOWITZ, J. C., MANDLER, G., & BARSALOU, L. W. Recognition failure: Another case of retrieval failure. *Journal of Verbal Learning and Verbal Behavior,* 1977, 16, 639–663.

RIPS, L. J., SHOBEN, E. J., & SMITH, E. E. Semantic distance and the verification of semantic relationships. *Journal of Verbal Learning and Verbal Behavior,* 1973, 12, 1–20.

RIPS, L. J., SMITH, E. E., & SHOBEN, E. J. Set-theoretic and network models reconsidered: A comment on Hollan's "Features and semantic memory." *Psychological Review,* 1975, 82, 156–157.

SMITH, E. E., SHOBEN, E. J., & RIPS, L. J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review,* 1974, 81, 214–241.

TULVING, E. *Elements of episodic memory.* London: Oxford Univ. Press, in press.

TULVING, E., & THOMSON, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review,* 1973, 80, 352–373.

TULVING, E., & WATKINS, O. C. Recognition failure of words with a single meaning. *Memory & Cognition,* 1977, 5, 513–522.

TULVING, E., & WISEMAN, S. Relation between recognition and recognition failure of recallable words. *Bulletin of the Psychonomic Society,* 1975, 6, 79–82.

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

WATKINS, M. J., & TULVING, E. Episodic memory: When recognition fails. *Journal of Experimental Psychology: General,* 1975, 104, 5–29.

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

(Received September 4, 1981)