

Implicit and explicit memory and response bias

J. ANTHONY DEUTSCH

University of California, San Diego, La Jolla, California

(John T. Wixted, *Sponsor*)

The influence of response bias on tests of implicit and explicit memory was assessed by using word-stem completion and recognition. During acquisition, subjects were presented either with a list of words unrelated to the task in the retrieval phase or with a list of words whose stems could be completed in only two ways (e.g. "appeal," "appear"). In the word-stem-completion retrieval phase, the subjects, who were given either explicit or implicit memory instructions, were asked to complete these stems. Retrieval of items in implicit and explicit memory tasks was highly correlated ($r = .593$ and $.625$, respectively) when the acquisition list had contained the words completing the stems and similarly correlated ($r = .667$) when the acquisition list had contained no words completing the stems. There was no correlation between recognition and explicit stem completion. In contrast to stem completion, there was no overall correlation between "recognition" in the condition in which the acquisition list contained words to be recognized later and the guessing condition, in which the acquisition list had contained no words presented during the retrieval phase. However, when the items were divided into two groups differing in memorability, a negative correlation appeared between recognition and guessing probability with the more memorable items, and a positive correlation appeared with the less memorable items. This suggests that response bias, rather than the memory the tasks are designed to test, is responsible for the normally reported lack of correlation between recognition and word completion.

In a widely used experimental paradigm for studying memory, subjects are presented with a list of words and then are tested, first using a recognition task and then using a cued-recall or word-completion task (for a review, see Nilsson, Law, & Tulving, 1988). A measure of statistical association is then computed between the scores of the two tasks. In the present experiments, a paradigm is used that measures the same essential relationships but that has been modified to enable a simple comparison between the way underlying biases operate on the two tasks between which a measure of association is calculated.

It has been proposed (Tulving & Schacter, 1990) that there are two independent types of memory, implicit and explicit, whose presence can be revealed by the use of different tests. The absence of correlation between the results of such different tests has been used to argue for the independence of two memory processes, although the validity of such arguments has been challenged (Hintzman & Hartry, 1990). However, the probabilities of response on the two tests could be influenced by another factor besides the memory processes being tested (e.g., item familiarity). This nonmemory factor or bias could be the major determinant of the pattern of results on one of the tests. If the major determinant of the other test being correlated is a different bias, or a memory factor, a lack of correlation between the tests would not show that there were two independent types of memory.

The three tasks used are two word-stem completion tasks and a recognition task. To render the stem-completion and recognition tasks comparable, words were used which, when truncated, had only one other completion in normal usage (e.g., "appear"- "appeal," "arcade"- "arcane"). Thus, the subjects, whether asked to complete a stem or recognize a word, are asked to choose between the same two alternatives, thus rendering the two different memory tasks quantitatively comparable. Furthermore, to assess the extent of baseline bias to choose one of the two alternatives, some groups of the subjects were asked to choose between the alternatives when they had not been exposed to the list of words containing either of the two alternatives, but had been given a list containing neither. (The results show that this method of assessing bias works well.)

In one task that was investigated, one group of subjects was presented with a list of words in the acquisition phase and then asked to complete a list of cues, consisting of truncated words, with the first word that came to mind. Such a word-completion task is similar to those used as tests for implicit memory (Tulving & Schacter, 1990).

In the second task that was investigated, another group of subjects was given the same treatment as the subjects in the first task, except with different instructions in the retrieval phase. Instead of being asked to complete a list of cues with the first word that came to mind, the subjects were instructed to complete the cues by using the words that appeared on the list in the acquisition phase. This second stem-completion task has been used as a test of explicit memory.

The third task explored here was a recognition task. In the acquisition phase, another group of subjects was given a list of words. (This consisted of the same words that were used in the previous two tasks.) In the retrieval

My thanks are due to my colleagues, E. B. Ebbesen and J. T. Wixted, for helpful discussion and to Melody Chen for help in statistical calculation. Correspondence should be addressed to the author at the Department of Psychology, University of California, San Diego, La Jolla, CA 92093-0109.

phase, pairs of words were presented, and the subjects were asked to indicate which one of each pair occurred on the list in the acquisition phase. Such a task is commonly used as a test of explicit memory.

EXPERIMENT 1

Method

Subjects. These were 68 undergraduates in a class on perception at the University of California, San Diego.

Materials and Procedures. These were three lists of 32 words. Two of the lists were selected in the following way. There are stems in the English language that can be completed in only two ways to form a word. As examples, there are the words "charity" and "chariot" and "appeal" and "appear." Such a stimulus set provides various methodological advantages. Thirty-two pairs of such words were selected, and some attempt was made to select stems that were pronounced the same way in their two completions. One word of each pair was placed on each of the two lists.

The third list again comprised 32 words, but these words were unrelated to the words on the first two lists. These three lists were shuffled and distributed to the subjects, with instructions at the top of each list to print their names and to rate the words for frequency of use on a 7-point scale. (The subjects were also asked to state if they were native speakers of English. Only the records of those that were used for analysis.) When the subjects had completed their ratings, the lists were collected. The whole class then listened to a lecture on somesthesia for 30 min. At that point, three other lists were distributed. Two of these contained the stems common to the words in the first two lists, along with 18 others that acted as fillers, making a total of 50 word stems to be completed. These first two lists were identical except for the instructions printed at the top of the lists. The implicit memory instructions were as follows:

Please PRINT your name at the top. Each of the following cues is the beginning of an English word. The number of dashes after each cue indicates the number of letters required to complete the word. Please complete each cue by using the first word that comes to mind.

The explicit memory instructions were as follows:

Please PRINT your name at the top. Each of the following cues is the beginning of an English word. The number of dashes after each cue indicates the number of letters required to complete the word. Please complete as many of the cues as you can by using the words that appeared on the first list you were given and were asked to rate for frequency of use.

The third list contained 64 words arranged in pairs. Each pair consisted of the two completions of each word stem (such as "charity"-"chariot," "edible"-"editor," "touch"-"tough," etc.). The instructions to the subject were as follows:

Please PRINT your name at the top. There are pairs of words on this sheet. One of each pair was on the list you were asked to rate. Please indicate by circling one word of each pair to show which one you think was on the list you were asked to rate.

As soon as the subjects completed their task, the lists were collected, and the lecture continued.

Results

The results are presented in Table 1. In an analysis of variance (ANOVA), the data were computed scoring all the

groups as if Relevant List 2 was correct. This could be done because the two relevant acquisition lists consisted of words that were the only alternative completions of a word stem. Thus, divergence from the score of the group presented with the irrelevant list leads to an increase in "errors" in groups presented with relevant list 1 whenever memory of this list is operating and a decrease in errors in the groups presented with relevant list 2. A 2×3 ANOVA (implicit vs. explicit instructions \times 2 relevant and 1 irrelevant list during acquisition) showed no main effect of implicit versus explicit instructions [$F(1,62) = 1.174, p = .283$], as would be expected. There should be no effect of implicit versus explicit instructions during retrieval, when subjects were essentially guessing (in the irrelevant acquisition list condition). Furthermore, the two groups that had been given the two complementary relevant acquisition lists should show the same difference from the guessing condition, but in different directions, so that these should cancel out when the mean was computed. We should, therefore, predict that the overall means of implicit and explicit instructions conditions would be highly similar, which was the case (explicit mean = 16.34, $SD = 5.6, n = 36$; implicit mean = 17.23, $SD = 3.7, n = 32$). On the other hand, a highly significant main effect of acquisition list was shown [$F(2,62) = 31.46, p < .0001$]. There was also a highly significant interaction between the two main effects [$F(2,62) = 5.63, p = .0006$], showing that the effect of the instructions during retrieval (implicit vs. explicit) depended on the type of list presented during acquisition. To assess the influence of memory on performance in the implicit retrieval condition, a comparison was made between the two pooled guessing conditions (implicit and explicit instructions when an irrelevant list had been presented) and the two pooled implicit memory conditions. To pool the memory conditions, errors were counted as failures to remember items on the list that had been presented to the subjects, with the number of "errors" for that list obtained in the guessing condition used as baseline. (This contrasts with the convention used in computing the ANOVA, in which errors were counted as successes for one of the groups.) The difference between means was significant [$t(44) = 2.41, p < .02$, two-tailed]. A further comparison was made between the two pooled implicit memory conditions and the two pooled explicit memory conditions. The difference between these means was also significant [$t(42) = 3.18, p < .003$, two-tailed]. There is a mean of about 14% of possible improvement in the two "relevant implicit" groups and of about 35% of possible improvement in the two "relevant explicit" groups. The recognition scores are much higher, a mean of only 1.7 er-

Table 1
Stem Completion Performance With Respect to "Errors" as a Function of Instructions During Retrieval and of List Presented During Acquisition

Instructions During Retrieval	List During Acquisition								
	Relevant List 1			Irrelevant List			Relevant List 2		
	<i>M</i>	<i>SEM</i>	<i>n</i>	<i>M</i>	<i>SEM</i>	<i>n</i>	<i>M</i>	<i>SEM</i>	<i>n</i>
Explicit	21.75	±1.11	10	17.57	±0.65	14	10.42	±1.3	12
Implicit	19.3	±1.05	12	17.15	±1.13	10	14.85	±0.96	

Note—Since the two relevant acquisition lists consisted of words that were the only alternative completions of a word stem, the data are presented scoring all the groups as if relevant list 2 was correct. Consequently, divergence from chance leads to an increase in errors (which reflect the memory of relevant list 1) in groups presented with relevant list 1 and a decrease in errors in groups presented with relevant list 2.

rors ($SEM \pm 1.2$) for the two "relevant acquisition list" groups. The mean for the irrelevant acquisition list could not meaningfully be computed, since this group almost unanimously declared that it had not seen any of the items.

Although there are differences in the mean number of errors between the implicit and explicit groups (in which a relevant list had been presented during acquisition), there are highly significant correlations between these memory conditions [$r(30) = 0.593$ and 0.625 , both $ps < .005$]. The relative probabilities of an error for any given item in the implicit and explicit groups are somewhat similar. Although the probabilities of any error are smaller in the explicit test, the relative probability of an error on a specific item in this group compared with the probability of an error on another specific item tends to be the same as in the implicit test. Thus, the results on the explicit and implicit tests are not independent.

However, the correlation between implicit and explicit tests when no memory is involved (when the subjects had been presented with an irrelevant list in the acquisition phase) equals .667. By using the χ^2 test for homogeneity of values of r (Edwards, 1984, p. 74) to test for the significance of differences between the three values of r (.667, .593, and .625), we do not obtain a significant result ($\chi^2 = 0.199$, 2 df , $p > .9$). Thus, the lack of independence between the implicit and explicit memory tests (when a relevant list had been presented in the acquisition phase) is not due to any significant extent to the operation of memory processes but to a response or guessing bias. The influence of a memory process or processes on the correlation between these memory tasks is inapparent.

EXPERIMENT 2

Method

As reported above, the subjects in Experiment 1 who were asked to recognize the words they had rated for frequency of use made very few mistakes. To make recognition performance comparable with that in explicit stem-completion tasks, other groups of subjects were run with an increased delay between initial learning and subsequent recognition test. The procedure was that used in Experiment 1. One group of subjects ($n = 17$, again undergraduates in class) was given the same irrelevant list of words that was used in the first experiment; another group of subjects ($n = 21$) in the same class was given a second list of words, the same as relevant list 1 in Experiment 1. Five days later, the subjects were given the same recognition list that was used in Experiment 1, consisting of pairs of words in which both members were the two completions of the same stem. The subjects who were asked to recognize the words that they had not seen completed their task satisfactorily. However, the subjects who were asked to recognize the words that they had actually been presented with still produced a much higher score than did the explicit-stem-completion subjects in Experiment 1. Accordingly, another group of subjects ($n = 18$, undergraduates from another class) was run using the same list of relevant words, but these words were presented to them again for recognition 21 days later. This time, the number of items correct was close to that obtained by the explicit-stem-completion group. An additional set of subjects (26-day group) who had been shown the same set of words as the 21-day group was run in the same recognition task 5 days later and was divided into two groups. The first group ($n = 30$) was simply asked to circle one of the words of each pair, one of which they were told was on the original list. The other group ($n = 41$) was asked only to circle the words that they were "quite sure" they recognized. The instructions, which were printed, were as follows:

Please PRINT your name at the top. There are pairs of words on this

sheet. Circle one of the words in the pair, but only if you are quite sure that it was on the list that you were asked to rate for frequency of use. If you are not sure, do not circle either of the words in the pair.

Results

The first set of results concerns correlations between the groups that had been shown an irrelevant list during the acquisition phase. They were then asked to perform either one of the stem-completion tasks (implicit or explicit) or the recognition task. The correlation between the recognition and stem completion in the explicit mode was .501 ($n = 32$, $p < .005$). The correlation between recognition and stem completion in the implicit mode was .591 ($n = 32$, $p < .005$); the correlation between the two types of stem completion cited above when the subjects were guessing was .667. The word in each pair that was thought to be recognized also tended to be rated as the more frequent one [as the correlation between putative recognition and ratio of frequency of use of the two words was .306 ($n = 32$, $p < .05$)]. This correlation is not significantly different from .47, obtained in the analogous condition in Experiment 1 ($p = .26$). Thus, the word that tends to be chosen from each pair during stem completion predicts to a significant extent the word that is chosen during a recognition task in which the subject has no memory of the correct alternative and is forced to guess.

In Experiment 1, we saw that there was a good correlation between the probability of guessing a word and the probability of remembering it correctly, both during stem completion. No such overall correlation is seen during recognition. The correlation between guessing recognition (as measured in the 5-day group) and "real" recognition in the 21-day group was .099 ($n = 32$, n.s.). Similarly, the correlation between guessing recognition (as measured in the 5-day group) and real recognition in the 26-day group was $-.12$ ($n = 32$, n.s.). This lack of correlation was surprising, since the mean number of items correct in the 21-day group was only 23.4 (22 for the 26-day group). This means that about half the items had been forgotten (the mean being close to 75% correct) and would consequently be guessed at. There should, therefore, have been some positive correlation between real recognition and guessing recognition—that is, unless those items that were still remembered were correlated in the negative direction with guessing and would thus cancel out the positive correlation that should exist for the forgotten items. On account of this line of reasoning, an attempt was made to determine which items were likely to have been remembered and which forgotten. As was described in the methods section when the recognition test was run on Day 26, an extra group ($n = 41$) was asked to circle only those items that they were "quite sure" of with respect to recognition. The number of subjects circling an item was counted to produce a score for each item. The scores for each of the items ranged from 32 to 2. Since about half the items were forgotten, the items were divided into two approximately equal groups. The group of items which was circled by 12 or more subjects contained 15 items. The second group of 17 items was circled by 12 or fewer subjects. The correlation between the real recognition and guessing recognition was computed separately for the more memorable and less memorable items. For the 21-day

group, there was a negative correlation of $-.285$ between guessing and recognition for the more memorable items but, for the less memorable items, a positive correlation of $.414$ between the guessing recognition task and the real recognition task ($n = 17, p < .05$; the probability of the difference between these negative and positive correlations is $.0309$). Similarly, for the 26-day group, there was a negative correlation of $-.45$ between guessing and recognition for the more memorable items ($n = 15, p < .05$) but a positive correlation of $.294$ between the guessing recognition task and the real recognition task for the less memorable items. (The probability of the difference between these negative and positive correlations is $.0224$.) Thus, both in the 21- and 26-day recognition groups, there is a large difference in the correlation between guessing and recognition of items when we compare more memorable and less memorable items. The recognition of less memorable items tends to be positively correlated with their probability when they are merely guessed, whereas the recognition of more memorable items tends to correlate in the opposite or negative direction with their probability of being guessed. The mean recognition percentage of correct recognition of the more memorable group of items tested at 21 days ($n = 15$) was 78.93 ($SEM \pm .90$), whereas the mean percentage of the less memorable group ($n = 17$) was 68.00 ($SEM \pm 3.59$). The difference was significant [$t(30) = 2.33, p = .0134$]. The mean percentage of correct recognition of the more memorable group of items tested at 26 days ($n = 15$) was 76.22 ($SEM \pm 2.88$), whereas the mean percentage of the less memorable group ($n = 17$) was 61.96 ($SEM \pm 2.53$). The difference was significant [$t(30) = 3.795, p = .0003$]. On the other hand, the difference between the means of the more and less memorable items when they are guessed is not significant. The more memorable items ($n = 15$) are guessed as correct 45.19% of the time ($SEM \pm 4.47$), whereas the less memorable are guessed as correct 51.2% of the time ($SEM \pm 5.12$). Such a difference would probably occur by chance [$t(30) = 0.856, p = .1994$]. Finally, there is the correlation between the two explicit tasks, recognition and stem completion (when subjects had been asked to use the list that had been presented). The correlation between recognition at 21 days and explicit stem completion was not significant ($r = -.134, n = 32$), and that between recognition at 26 days and explicit stem completion was similarly nonsignificant ($.019, n = 32$). This is in marked contrast to the good correlation between implicit and explicit stem completion reported above (Experiment 1).

DISCUSSION

Experiment 1 shows a good correlation between tests of implicit and explicit memory using stem completion. As the size of the correlation is about the same ($r = .63$) whether the subjects are guessing the answers on the tests or whether they had been presented with the relevant items in the acquisition phase, the correlation is not due to the memory component of the two tasks but rather to a common response bias or guessing strategy. The correlation obtained between these tasks is therefore not relevant to questions regarding the dependence or independence of memory processes underlying these tasks. The response bias seems due to the greater probability of selecting a more familiar item, since there is a correlation between the perceived frequency of use of a word and its selection in a stem-completion task. A second conclusion that can be drawn from Experiment 1 is that neither of the two sets of relative prob-

abilities compared in the implicit and explicit stem-completion tasks are determined to any significant extent by memory, but rather by a response bias or guessing strategy. This means that a lack of correlation between the results of a stem-completion task and the results of another task that reflects memory factors to an unknown degree cannot validly be used to postulate the presence of two independent memory processes, as has often been the case. There may be two independent processes, although not necessarily so, as will soon be shown. However, the independence that may exist need not be between two memory processes but between a memory process and a nonmemory process.

Experiment 2 found no overall correlation between response bias (or guessing strategy) and recognition. However, the experiment showed that this overall lack of correlation was not due to independence between the two sets of probabilities being compared but to the presence of two subsets of items in the recognition set, one positively and the other negatively correlated with response bias or guessing strategy. Items that are remembered with greater certainty correlate negatively with guessing; those that are remembered with lower certainty correlate positively with guessing. (The probability of guessing an item correlates positively with its perceived frequency of use, hence familiarity.)

Since response bias or guessing strategy correlates well with stem completion, the results of Experiment 2 throw a further doubt on the prevailing interpretation of the lack of correlation found between stem completion and recognition. Given the results of Experiment 2, it seems plausible that the lack of correlation between stem completion and recognition is not due to the independence of the two tasks. Instead, it seems that the probability of an item being selected both in recognition and in stem completion is determined by its familiarity. In stem completion, familiarity is positively correlated with the selection of an item whether memory for the item is present or absent. In recognition, familiarity is positively correlated with the selection of an item only if memory for the item is absent. But in recognition, when memory for an item is present, familiarity of an item is negatively correlated with its selection. Thus, when recognition of a list of items is correlated with its stem completion, the remembered items will be negatively correlated and the guessed at items positively correlated, thus tending to produce an appearance of independence between the two tasks.

The present results are to a great extent predictable from the large body of work on the word-frequency effect in recognition and recall. It has been found that rare, low-frequency words are better recognized than common, high-frequency words. On the other hand, Glanzer and Bowles (1976) found that when neither of the words to be recognized had been presented, the common, high-frequency words tended to be chosen. (When both of the words to be recognized had been presented, the rarer words tended to be chosen, as would be expected from the research quoted above.)

A second set of studies of relevance to the present analysis concerns the relationship of recall to word frequency. In these studies, there is a positive correlation between recall and word frequency, as in both recognition, in which subjects are guessing, and the reverse of recognition, in which memory is involved.

The results of the studies on the word-frequency effect predict the data of Experiment 2 and explain the lack of correlation between recognition and cued recall (which has been used as evidence for two independent memory processes), given some minimal assumptions.

REFERENCES

- EDWARDS, A. L. (1984). *An introduction to linear regression and correlation* (2nd ed.). New York: W. H. Freeman.
- GLANZER, M., & BOWLES, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning & Memory*, *2*, 21-31.
- HINTZMAN, D. L., & HARTY, A. L. (1990). Item effects in recognition and fragment completion: Contingency relations vary for different subsets of words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 955-969.
- 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.
- TULVING, E., & SCHACTER, D. L. (1990). Priming and human memory systems. *Science*, *247*, 301-306.