Interference paradigms for meaningful propositional memory

Gordon H. Bower
Stanford University

Interference phenomena in the learning, transfer, and retention of meaningful, factual propositions were investigated. Subjects first learned 20 sentences asserting a fact about a character denoted by a profession name, e.g., "The lawyer paid his insurance." Recall was cued by giving the sentence-subject (A-term) for recollection of the verb phrase (B term of A-B sentence). Different groups then received one of four different interpolated tasks: new predicates (A-D), reshuffled predicates (A-Br); new professions and predicates (C-D), or no learning (Rest control). Retention of the initial A-B sentences was tested by fast-paced forward recall, then slow-paced full recall of everything to the stimulus ("MMFR" test), then "backward" recall, then pair-recognition memory. Interference theory correctly predicted all results. A-Br showed greater negative transfer than did A-D relative to C-D. Forward retention was best for the Rest group, then C-D, then A-D, and A-Br worst. Backward recall and recognition memory were equally high for the Rest, C-D, and A-D conditions, and poor for the A-Br condition. It was concluded that interference theory applies as well to meaningful propositional learning with gist recall as it does to verbatim memorizing in standard rote verbal-learning tasks.

This experiment was undertaken to determine whether the learning, transfer, and retention of meaningful sentences can be partially described by the laws of associative interference. The answer is not a foregone conclusion. Many psychologists believe that associative interference is a phenomenon appropriate only to verbatim reproduction of usually nonsensical materials. But I would hope to show that interference applies just as readily at the level of forgetting the "meaningful gist" of sentences. However, surprisingly few experiments have attempted to prove that point.

A few prior experiments have demonstrated retroactive interference of declarative sentences in one specific paradigm reminiscent of the A-B, A-D paradigm of paired-associates. Anderson and Carter (1972) had people learn two lists of sentences where the subjects of the
sentences were repeated with new predicates, e.g., “The nurse walked her dog” was in List 1 and “The nurse helped the doctor” was in List 2. They found that learning of the second list of sentences increased forgetting of the first list when tested by cueing with the sentence subject. The forgetting of the A-D group was larger than that of control subjects who learned a second list of sentences with all new sentence subjects and predicates; this was the equivalent of an A-B, C-D control condition.

Anderson and Bower (1973) also found negative transfer and retroactive interference in a multilist sentence-learning experiment where sentence subjects were repeated across lists with changing predicates. In addition, they showed that reaction time to verify that a particular subject-predicate sentence had been learned increased the greater the number of unrelated predicates linked to the same subject, and also increased the greater the number of different subjects linked to the predicate of the test sentence. This interference in verification latencies has been repeatedly found (e.g., Anderson, 1976).

If we think of the subject of a simple declarative sentence as the “A-stimulus” and the predicate of the sentence as the “B-response,” then a sentence can be thought of analogously to an A-B paired associate that the person learns. In these terms, the interference paradigms explored so far with sentences are A-B, A-D and A-B, C-D, and the primary effect studied has been poorer recall of A-B following interpolated learning. But classical interference has a much larger set of paradigms that have been systematically explored; moreover, many interesting patterns of performance can be assessed if we employ other memory measures, such as modified free recall, backward recall, list differentiation, and recognition tests. The issue in the present experiment was to explore with sentence learning more of the classical interference paradigms than had been done heretofore and to assess memory and interference in several different ways.

To this end, we trained four different groups of subjects on three successive lists of declarative sentences whose subject-predicate pairing exemplified four different paradigms: A-B, Rest; A-B, A-D, A-F; A-B, C-D, E-F; and A-B, A-Br, A-Br’. The notation A-Br means that the second list contains the same predicates and subjects as the first list but their pairing differs. Henceforth, I will abbreviate these four groups as Rest, A-D, C-D, and A-Br. Two interpolated lists were used to boost the amount of interference operating on recall of List 1. Each group was tested for memory of A-B in four different ways. First, forward recall (recall B to cue A) was tested at a very fast pace; performance scores here will be primarily sensitive to long reaction
times created by interference. Second, forward recall was tested in a slow-paced modified free-recall ("MMFR") procedure wherein the person tried to recall all the predicates associated to a given sentence-subject. Performance scores here should be sensitive to unlearning of initial associations caused by interpolated learning (see Barnes & Underwood, 1959). Third, backward recall (recall A to cue B) was tested, and this should reveal strong interfering effects only for the A-Br condition. Fourth, multiple-choice recognition of A-B was tested, with a sentence subject (A term) listed beside five predicates (B terms) from the first list, of which one was correct, two had been paired with that stimulus term in Lists 2 and 3 for the A-Br condition, and two were completely incorrect. On this recognition test, we expected performance to be especially poor for the A-Br condition but not for the others.

We can also examine transfer effects on second- and third-list learning in this design. In particular, we expect A-D to show slower learning ("negative transfer") with respect to the C-D condition. The A-Br group should show greatest negative transfer of all.

In the retention tests, overall A-Br should be worst, then A-D, then C-D, and the Rest group should do best. The C-D condition should do worse than the Rest group in recall on the basis of "generalized response competition" or "response-set suppression" (see Postman & Underwood, 1973). The A-D condition should have poorer forward recall than the C-D condition because of unlearning of the forward association (A-B) in the A-D condition. Conditions Rest, C-D, and A-D should not differ in backward recall because they have learned no competing association to the predicate term, B, which is the cue for recall. On the other hand, condition A-Br should do poorly on backward recall. In recognition memory of the A-B pairings, conditions Rest, A-D, and C-D should perform at a high level because response-set suppression is immaterial to such a recognition test and because pair recognition can be effected on the basis of an intact forward or backward A-B association. Only A-Br should do poorly on the pair-recognition test; that is because during interpolated A-Br learning both forward and backward A-B associations are being unlearned, so neither will be available to support pair recognition. The recognition problem was compounded for the A-Br subjects since two of the predicate lures on the multiple-choice test were the predicates that had been associated to the A sentence-subject in Lists 2 and 3. Thus, in trying to remember which predicate was associated with the A subject in List 1, the person has to overcome a list discrimination problem. The two factors—unlearning and list discrimination difficulties—
should combine to produce poor performance of the A-Br group on the pair recognition test.

The following experiment was carried out to test these interlocking predictions.

**METHOD**

**Subjects**

The 48 subjects were Stanford undergraduates, half of them men and half women. Twenty-four of the subjects were recruited by an ad in the campus newspaper and were paid for their participation. Twenty-four subjects participated to fulfill a service requirement for their introductory psychology course. The subjects were tested individually, assigned in random alternation in cycles of four to the four experimental conditions. Thus, each experimental group had 12 subjects.

**Materials**

For generality we constructed two different versions of List 1, the “A-B” list. The lists were composed of 20 unrelated sentences. Each was a simple declarative sentence of the form “The Profession-Name Verbed Noun-Phrase.” Examples are “The fireman watered his plants,” “The teacher sold his house,” and “The carpenter shot a deer.” Twenty different, common professions and verb phrases were used for Version 1 and for Version 2. Half the subjects in each condition learned Version 1 for A-B, and half learned Version 2. The versions were constructed to be of comparable difficulty, and the results proved them to be so.

Appropriate interpolated lists were composed in relation to Versions 1 and 2 of the first, A-B list. For A-D, A-F, the 20 profession names of List 1 were retained but they were paired with new verb phrases for Lists 2 and 3. For example, if List 1 had the sentence “The fireman watered his plants,” List 2 might have “The fireman gave to charity” and List 2 have “The fireman envied his brother.” An effort was made to keep the successive verb phrases attached to a subject semantically independent.

The A-Br and A-Br’ lists were composed by randomly re-pairing the professions and predicates of the first first, insuring that no pairing was duplicated across lists. The C-D, E-F lists were composed by pairing 20 new profession names per list with the new D and F predicates used in Lists 2 and 3 of the A-D condition. The C-D list for Version 1 was in fact the A-B list for Version 2, and vice versa.

Four different test materials were constructed. For the “fast forward” test, the profession names (A-terms) of the person’s first list were printed on a memory-drum tape and presented at a 2-sec-per-item rate in the window of a memory drum. In order to accommodate the subject to the rapid pace of the memory drum, the profession names were preceded by six single digits to which the person was instructed to add 3 and say the sum aloud within the 2-sec memory-drum cycle, e.g., presented with 4, he or she was to say “Seven.” Then the words “Ready to Test” were followed by the first profession name, for which the person was to recall the appropriate verb phrase from List 1. The other profession names followed at the 2-sec pace.
The MMFR test was on a sheet of paper the subject filled out, having 8 min. to do so. The MMFR sheet essentially tested the subject's forward recall (from profession cues) of every sentence learned in Lists 1, 2, and 3. It was emphasized that we were particularly interested in the subjects' recall of List 1. The profession-name cues were listed down the left side of the recall sheet with either one blank space to the right (for the Rest condition) or three blank spaces to the right, marked "List 1, List 2, List 3" (for the A-C and A-Br conditions). For the C-D condition, the 20 profession names from each list were listed in three columns (titled Lists 1, 2, 3) with a blank space to the right of each, and the person was to try to recall all 60 verb phrases, beginning with List 1.

The backward-recall test was composed of the 20 verb phrases from List 1 written on a sheet with a blank space to the left into which the subjects were to write the profession name that had been paired with that verb phrase in List 1. They had 3 min. to do so.

The final test was a five-alternative multiple-choice test of the first, A-B list. The 20 profession names were listed down the column of the page, and five predicates from List 1 were printed to the right of each profession. The person was to put a "1" beside his best guess at the predicate paired with that profession name in List 1, and put a "2" beside his second-best guess. The multiple-choice alternatives were constructed appropriately to the A-Br condition; that is, two of the choices were verb phrases (of List 1) which the A-Br person had learned for that profession in Lists 2 and 3.

Procedure

The subjects were tested individually, seated at a lab table across from the experimenter. The lists were learned by the anticipation method using flash cards with the profession (as cue) written on one side and the full sentence written on the other. Each list began with a study trial at a 5-sec rate per sentence. Then the items were tested. The profession name was shown for 5 sec while the subject tried to recall the verb phrase; after that the sentence was shown for study for 5 sec. For the critical List 1, in order to keep degree of learning at a low level and promote interference effects, each item was dropped out of the training cycle when it was anticipated correctly for the first time; on that occasion, only the feedback "Right" was given to the person's recall and he or she was not shown the full sentence again. List 1 training was continued by the drop-out method until each item had been anticipated correctly once. A consequence of the drop-out procedure, of course, is that the effective list-length decreases as progressively more items are dropped from the pool of to-be-learned items. The remaining cards were shuffled after each trial. After all List 1 sentences had been recalled once correctly, the subject was given a 1-min. rest and instructions for learning List 2. There followed a study trial and then three anticipation test-study trials on the sentences of List 2. The test and study intervals were 5 sec each. All 20 sentences were run all trials without drop-outs. Following List 2 another 1-min. rest ensued and then List 3 was learned by the same procedure of four study trials and three tests without dropouts.

Interpolated learning of the two lists required 20 min. Subjects in the Rest control condition were engaged in an interesting distraction for these 20 min., namely, rating the "funniness" of a set of "Peanuts" cartoons on a 1 to 4 scale.
The retention tests began after the interpolated learning tasks (or the Rest distractions) were completed. The majority of subjects spontaneously expressed surprise that retention of earlier lists was to be tested. The nature of each retention test was carefully explained before it was done. The “fast forward” test on the first list was carried out first on the memory drum after accommodating the person to the 2-sec rate at which he or she would have to recall the verb phrase to the profession cue. Then the 8-min. MMFR recall sheet was filled out by the subject. We had emphasized that recall of List 1 verb phrases was of most importance to us and so those should be done first. The subject was told to adopt a lenient recall criterion, to guess if only mildly sure of his answer, since we were interested in his guesses. The backward-recall test and the multiple-choice test followed.

All subjects received all tests in the order specified. Though there are undoubtedly influences of the early tests upon later ones, the order of the specific tests was designed to minimize such influences. For example, the backward test, which should enhance the availability of the verb phrases, was not given until after the forward-recall tests, which assess verb-phrase availability; the slow-forward test, which should enhance the availability of forward associates, was not given until after the fast-forward test, where performance depends on that availability. It was felt that the valuable information obtainable from the multiple memory tests more than offset the possible distortions in test results introduced by test-sequence influences. In any event, these test sequential influences should not be different across the four experimental conditions.

The verb-phrase recalls were scored by a lenient “gist” criterion, giving half a point for the verb and half for the noun phrase. It was rare for the person to recall one without the other.

RESULTS

Original learning

The four groups were comparable in rate of initial learning of A-B sentences, with mean trial number of the first correct response varying insignificantly from 2.6 to 3.1 for the four groups. Also, Versions 1 and 2 were of equivalent learning difficulty. They were also equivalent in retention in all respects; henceforth we will pool the results from subjects learning the two different list versions.

Negative transfer

Figure 1 presents the percentage of correct responses over the three trials on List 2 for the C-D, A-D, and A-Br conditions, respectively. For each trial, the groups are ordered as listed in percentage recall. Assuming that the C-D condition serves as a control for learning rate, the A-D and A-Br conditions clearly show negative transfer relative to C-D. Averaging over the three trials, the mean percentages of correct recalls were 84 for C-D, 75 for A-C, and 63 for A-Br, and these
INTERFERENCE AND MEMORY

Figure 1. Percentage of correct recalls over three trials on List 2 for the C-D, A-D, and A-Br conditions
differed reliably in the expected direction ($p < .01$). In learning List 3 the group averages lined up in the same order, but the differences were too small to be significant.

So, we may conclude that the interlist relations produced negative transfer of the expected orders, at least for List 2 learning. By List 3, subjects had become such proficient learners that there was little chance for negative factors to show themselves.

Retention measures

The retention measures for the first list sentences are summarized as recall percentages in Table 1. There are clearly wide variations in performance as a function of experimental condition and testing procedure. We will proceed through the results from left-to-right in Table 1.

Table 1. Retention measures for the four groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Fast forward</th>
<th>Slow forward</th>
<th>Backward recall</th>
<th>Recognition test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>L1</td>
<td>L2</td>
<td>L3</td>
</tr>
<tr>
<td>Rest</td>
<td>63</td>
<td>83</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C-D</td>
<td>49</td>
<td>70</td>
<td>91</td>
<td>93</td>
</tr>
<tr>
<td>A-D</td>
<td>22</td>
<td>54</td>
<td>68</td>
<td>90</td>
</tr>
<tr>
<td>A-Br</td>
<td>19</td>
<td>40</td>
<td>40</td>
<td>80</td>
</tr>
</tbody>
</table>
**Fast-forward test**

As expected, the fast-forward test was the most difficult test of all. The four conditions differed significantly in all overall test, \( F(3, 44) = 15.2, p < .001 \). There was a statistically significant gap between percentages of the Rest and C-D groups, and between the C-D and A-D groups, but the A-Br was not significantly worse than the A-D group at this point.

**MMFR test**

Recall of each of the three lists on the MMFR test is shown in Table 1. Focusing first on recall of List 1, the Rest condition is clearly best, then the C-D, A-D, and A-Br groups in that order. This is the predicted order. The overall differences are significant, \( F(3, 44) = 12.9, p < .001 \). There is a significant gap between the C-D and A-D conditions \((p < .05)\), and so other group comparisons falling across this gap are significant. The List 1 recalls on the MMFR significantly exceeded recall on the fast-forward test for each condition.

The recalls for List 2 also fell in the predicted order. List 2 recall was higher for C-D than for A-D, \( F(1, 22) = 17.8, p < .001 \), and was higher for A-D than for A-Br, \( F(1, 22) = 16.3, p < .001 \). Similarly, the groups performed in the predicted order for recall of List 3, though the differences were not great. Recall across the three lists uniformly showed List 1 lowest, then List 2, then List 3 highest. This ordering is expected from interference theory, although in the present case input order is also confounded with retention order; also, List 1 was learned to a criterion different from those in Lists 2 and 3, so the three lists are not strictly comparable in degree of learning.

**Backward recall**

It was predicted that backward recall, that of the profession name when cued with the verb phrase of List 1, would be equally high for all groups except for the A-Br group. The results in Table 1 confirm this prediction. The Rest, C-D, and A-D groups do not differ from one another, but they each exceed the A-Br group, all \( ps < .01 \). Interestingly, the backward recall exceeds the List 1 forward recall (on the MMFR test) for all the conditions, significantly so for the A-D and C-D groups. This ordering probably reflects the priming of the “profession names” as available responses by the subject having just seen those professions on the forward tests that preceded the backward tests.

**Recognition test**

For the recognition test, we counted as a “hit” only cases where the correct answer was rated the “first choice.” High performance was
expected in all groups except for A-Br, and it occurred. The Rest, C-D, and A-D groups did not differ, but all were superior to the A-Br group (all \( p < .001 \)). A large percentage (70%) of the errors in the A-Br condition arose from subjects selecting the predicate paired with the profession name on List 2 or 3 rather than List 1. This indexes the problem of correct list identification that was mentioned earlier.

**Intrusion errors**

Intrusion errors can sometimes be informative regarding sources of interference; they depend, however, upon the person's subjective criterion of when overt guesses are appropriate. We examined interlist intrusions, specifically, the importation into recall of List 1 verb phrases that were appropriate to List 2 or 3 that the person had studied. On the fast-forward test, the mean number of other-list intrusions per subject was 0.4, 1.6, 2.1 for the C-D, A-D, and A-Br groups, respectively. These fall in the expected order and differ significantly, overall \( F (2, 33) = 4.91, p < .01 \). Similarly considering attempts to recall List 1, on the MMFR test, interlist intrusions averaged 0.08, 1.25, and 1.75 per subject for the C-D, A-D, and A-Br groups, respectively. Again these fall in the predicted order, though numbers and differences are quite small.

**DISCUSSION**

It appears that nearly every prediction of interference theory for this experiment has been confirmed. We have found appropriately graded amounts of negative transfer and retroactive interference, and the experimental groups performed in the patterns predicted on the various memory tests. For example, the C-D group fell between the Rest and A-D groups on the forward-recall tests, but all were equal on the backward-recall and recognition memory tests. Beyond the correct performance patterns, the intrusion errors on the retention tests came out in the predicted order, with most for A-Br, then A-C, and then C-D least. All in all, the results suggest that interference theory and interference paradigms apply just as readily to meaningful propositions as they do to the verbatim learning in standard verbal-learning tasks.

A critic might complain that we have demonstrated interference phenomena in sentence learning by the simple device of forcing the subject to treat it as a rote-learning task. But that is simply untrue, from both the experimenter's and subjects' viewpoints. The subjects were told to imagine that they were learning about a number of characters in a small town, much the way they might when reading a
story. They were told to use whatever imaginative locations, persons, and imagery they wished to learn the material, and most reported meaningful elaboration of the learning sentences. Also, gist recall, not verbatim recall, was emphasized, and subjects received feedback on the correctness of their gist recall. However, we were surprised that subjects overwhelmingly recalled close to the verbatim correct phrases rather than paraphrases—a result that Kintsch and Bates (1977) have also reported for subjects’ recognition memory for statements in a lecture they had heard. Further, there is nothing the least artificial about learning these facts about hypothetical characters, certainly no more so than a student’s learning facts about real historical characters. There is every reason to believe that we would have obtained exactly the same results if we taught the subjects some biographical facts about actual but somewhat unfamiliar historical characters (e.g., generals of the Confederate army at the outbreak of the Civil War).

The point of these remarks is to defend the idea that interference theory can be used to predict transfer and retention of meaningful sentences, of the conceptual knowledge that such sentences express. Although ideas from psycholinguistics are needed to understand relationships among sentences and the knowledge they express, when we turn to questions concerning transfer and retention of propositional learning, it seems we must fall back upon interference theory and look for some interpretation of the linguistic materials that permits clear application of interference concepts. In the present case, the part of the sentence used to cue its recall (the grammatical subject) was construed to be the “stimulus” or “A term” of an A-B association, and the verb phrase of the sentence was construed to be the “response” term. With those identifications, the interference theory was easily applied and the experimental paradigms were simply constructed. The results, it must be conceded, were a resounding success for the theory.

Notes
This research was supported by research grant MH-13905 to the author from the National Institute of Mental Health. Thanks are due to Marilyn Broyhill Beech, who assisted in the experiment. Requests for reprints should be sent to Dr. Gordon Bower, Department of Psychology, Stanford University, Stanford, CA 94305. Received for publication March 6, 1978.

References


