Failure to replicate mood-dependent retrieval

GORDON H. BOWER and JOHN D. MAYER
Stanford University, Stanford, California

We tried to replicate mood-dependent memory retrieval in a two-list interference design, as reported earlier in Experiment 3 of Bower, Monteiro, and Gilligan (1978, "Emotional mood as a context for learning and recall," *Journal of Verbal Learning and Verbal Behavior*, **17**, 573-585). Subjects in the experimental condition learned two word lists for free recall, one list while feeling happy, a second list while feeling sad. Later they recalled both lists while feeling happy or sad. Results revealed no mood dependency in recall; subjects recalled about the same percentage of words from the same-mood list as from the opposite-mood list. Thus, mood-dependent retrieval must be judged to be an unreliable phenomenon in the laboratory.

Memory retrieval is said to be mood-dependent if it varies according to the similarity between the subject's mood at learning and at retrieval. Bower, Monteiro, and Gilligan (1978) reported three attempts to produce mood-dependent retrieval (henceforth MDR). In the first two experiments, subjects learned a *single* word list while feeling happy or sad; later, some subjects recalled the list while feeling happy, others while feeling sad. In neither experiment was MDR found; subjects recalled about the same when tested in the same as in the opposite mood. Other experiments using one-list designs have found either no MDR (Isen, Shalker, Clark, & Karp, 1978) or only asymmetric MDR (Leight & Ellis, 1981, Experiment 2; Macht, Spear, & Levis, 1977).

In the third experiment of the Bower et al. (1978) report,[1] a memory interference design was used in which each subject learned *two* word lists, one while happy, the other while sad. Later, subjects recalled both lists while feeling happy or sad. These subjects appeared to show mood-dependent retrieval. Recall percentages averaged 70% when the learning and testing moods matched and 46% when they mismatched. Thus, Bower et al. concluded that MDR occurred in interfering circumstances, which resembled ecologically valid conditions of multiple learnings in multiple moods and contexts.

MDR was also found in a subsequent undergraduate Honors thesis by B. Thompson, directed by Bower, and a study by Bartlett and Santrock (1979) (both subsequently reviewed in Bower, 1981). MDR for the two-list interference design was later reported by Bartlett, Burleson, and Santrock (1982), by Schare, Lisman, and Spear (1984), and by Goerss and Miller (1983).

Disturbing results, however, come from two recent failures to obtain MDR using the two-list interference design. Anne Ellis (1983) found no MDR in the two-list design when the happy and sad moods were induced by having subjects view a funny or a sad movie segment. In a partial replication of the two-list interference experiment,

The authors' mailing address is: Department of Psychology, Stanford University, Stanford, CA 94305.

Betty Bayer (1982, personal communication), working in Professor Spanos's lab at Carlton University, observed no MDR when she used hypnotic moods.

These doubts caused us to attempt a replication of Experiment 3 of the Bower et al. (1978) report, with only minor variations in detail. The variations in the word lists and the tasks interpolated between learning and delayed retention should not influence MDR.

METHOD

Subjects
Student volunteers, who responded to campus ads or who were enrolled in introductory psychology were screened for hypnotic ability on the Harvard Group Scale of Hypnotic Susceptibility (Shor & Orne, 1962). All subjects used (N = 48) scored 6 or higher on the Harvard scale (mean = 9.13); they were paid $5.00 or given course credit for participating in the experiment, which took about 90 min.

Word Lists
The words learned were 16 pleasant and 16 unpleasant nouns with concreteness ratings of less than 3.0 on a 7-point scale (Paivio, Yuille, & Madigan, 1968). Seventeen of the 32 words were used in Experiment 3 by Bower et al. (1978). Replacements eliminated unfamiliar words or clarified their emotional content. Eight pleasant and eight unpleasant words were assigned to each list.

Design
Our design mimicked that of the original. During the first mood, subjects learned a first list by free recall; during the second mood, they learned the second list. After a filled interval, a third mood, which matched the first or the second mood, was induced and the subjects were asked to recall all words from the first list. After finishing List 1 recall, they then recalled List 2. During this final mood, a recognition memory test for both lists was also administered. List order during learning was counterbalanced across subjects and conditions.

The three moods the subjects received depended upon condition. Control subjects learned both lists and recalled them all in the same mood, using different imagined scenes to create the three mood instances. This condition is symbolized by AAA, to indicate that the mood was of Type A during all three mood periods. The second condition is ABA; these subjects had a mood reversal (from happy to sad, or vice versa) for the learning of the second list, and then reverted to their initial mood for final testing. The third condition is symbolized by ABB, indicating that the subjects' mood was reversed for learning of the second list and was reinstated as the third mood for the final retention tests. Half the subjects in each condition started with a happy mood, and half with a sad

mood, making six mood conditions in all. Whenever a mood was repeated for a subject, a different emotional scene from the past was used to induce it.

Procedure

After the screening, hypnotizable subjects were telephoned, asked to participate, and told of the mood inductions. If they agreed, they were scheduled for the main experiment.

The subjects were tested individually in an office converted for hypnosis. The subjects sat comfortably in a reclining chair, lights were dimmed, and hypnosis was induced by the eye-closure technique (see instructions in hypnotic test Form C, Weitzenhoffer & Hilgard, 1962). The appropriate mood was then induced by asking the subjects to recall a happy or a sad event from their lives, to experience the emotion of that recollection, and to adjust its intensity until it was intense but not overwhelming. The subjects typically reported the emotion within about 1 min. They were then told that they would be able to maintain the emotion at this intensity throughout learning the word list.

After inducing this first mood, the experimenter read a 16-word list aloud twice, in a different random order each time, at 3 sec/word. The subjects were then asked to recall orally as many of the words as they could, in any order, for 2 min. After recall, the list was read a third time in another random order, and the subjects again had 2 min for oral free recall. The subjects then rated their mood on a 10-point scale, after which the first mood was removed. The second mood was induced, and the same procedure was repeated for List 2. After this, a posthypnotic suggestion for easy reentry into hypnosis was given, and then trance was removed.

Upon awakening, the participants were questioned; they reported no serious discomfort from the trance. A short break was provided, if desired. During the remainder of the 10-min period after awakening, letter-cancellation tasks were administered. After 10 min, the subjects were instructed to reenter trance and to signal the experimenter when they had done so. After the signal, trance-deepening instructions were given, and the final mood was induced by asking the participants to recall a happy or a sad life event different from the memory that they had used to induce the earlier similar mood.

Once in the third mood, the subjects had 3 min to recall orally List 1 and, immediately afterwards, 3 min to recall List 2. This recall was followed by a third mood rating. If mood had weakened by this time in testing, the subjects were asked to reintensify the mood. The recognition test followed, with mood refreshers given halfway and three-quarters through the test. The recognition memory test consisted of 64 words: the 16 words from each learned list and 32 distractors, selected from a similar word pool. The subjects' task was to identify each word as having been either on the first, second, or neither list that they had learned.

The subjects were then awakened and filled out a brief questionnaire about the moods they had experienced and their present mood. The experimenter discussed with them the hypnotic procedure, but they were not told the purpose of the experiment. After ensuring their normal mood, and answering any questions, we excused them from the experiment.

RESULTS

Dropouts

All but two students contacted agreed to participate. Four of the 52 participants were not used: one due to an experimenter error, one for exhibiting total hypnotic amnesia, one for falling asleep, and one who appeared aware of earlier MDR findings and may have altered his performance accordingly.

Mood Manipulations

Moods were verified in two ways. First, the subjects rated their mood's intensity after each learning and recall period on a 10-point scale, where 6 or 7 was very intense and 10 was overwhelming. Average intensity was adequately maintained at ratings of 5.99, 6.58, and 5.94 for the successive moods within the session. Second, MDR requires that mood during retrieval be similar to the previous mood of the same type and different from the contrasting mood. All but the first six subjects rated similarity among the three moods on a postexperimental questionnaire on a 7-point scale (from 1, very different, to 7, identical). For subjects in the ABA and ABB conditions, the mean similarity rating between their recall mood and the similar learning mood was 5.04, and that between their recall mood and the different learning mood was 2.00. Thus, the mood similarity ratings differed appropriately [$t(28) = 7.48, p < .0001$].

Next, we checked whether the two word sets that served as Lists 1 and 2 were equivalent in learning. Overall original recall for Word Sets A and B were nearly identical at 10.27 and 10.19 words, respectively. Thus, the word-set factor will be ignored.

Mood-Dependent Retrieval

Table 1 reports three measures of retention. The first, percent retention, is the final correct recall (at the retention test) for a given list divided by the number of words correctly recalled at the end of original learning on that list. The second measure, recognition, was calculated for List 1 as the probability that when the subject judged "List 1," the test item was from List 1, minus the probability that it came from the other list or the distractor pool. This measure was calculated similarly for recognition of List 2. The third measure, recall composite, is described below.

For MDR to occur, the subjects in the AAA and ABA conditions must recall more of List 1 than the subjects in the ABB condition. As Table 1 shows, the average recall percentages for List 1 across the three conditions do not differ. Although the recognition data did show differences between conditions, these differences were in the wrong direction, with ABB outperforming ABA. Table 1 shows results of a one-way ANOVA on arcsin transforms for each of the measures.

We next analyzed recall retention percentages for List 2, for which according to MDR, subjects in the ABB and AAA conditions should recall more than those in the ABA condition. Table 1 again shows no mean differences in the List 2 recall percentages. As before, ABB exceeds ABA in recognition memory, and List 2 is consistently better than List 1.

As a second analysis of MDR, the subjects should better recall the list learned in the same mood than they recall the list learned in the opposite mood. Difference scores were derived in the ABB and ABA groups by subtracting subjects' percentage recall of List 1 from their percentage recall of List 2. According to MDR, the subjects in the ABB and ABA conditions should have positive and negative differences, respectively. However, difference scores did not differ reliably.

Several similar analyses with different dependent measures were conducted. These included several alternative measures in the numerator of the recall ratio. One numer-

Table 1
Group Means and Standard Deviations (SDs) for Retention of Lists 1 and 2, and F Tests for Group Differences

	AAA		ABA		ABB		
	Mean	SD	Mean	SD	Mean	SD	F Test
List 1 Retention							
Percent Retention	.40	.18	.49	.21	.53	.21	$F(2,45) = 1.71, p < .19.$
Recognition	.53	.22	.40	.13	.60	.19	$F(2,45) = 5.15, p < .01.$
Recall Composite	.49	.19	.73	.35	.69	.20	$F(2,45) = 4.68, p < .01.$
List 2 Retention							
Percent Retention	.71	.18	.58	.16	.67	.21	$F(2,45) = 2.05, p < .14.$
Recognition	.55	.23	.47	.17	.61	.18	$F(2,45) = 2.05, p < .14.$
Recall Composite	.87	.25	.68	.26	.90	.28	$F(2,45) = 5.18, p < .01.$
Difference Scores (List 1 − List 2)							
Percent Retention			−.10	.30	−.14	.27	$F(1,30) = 0.32, p < .57.$
Recognition			−.07	.12	−.01	.11	$F(1,30) = 3.12, p < .09.$
Recall Composite			.05	.41	−.22	.34	$F(1,30) = 6.03, p < .02.$

ator was the correct list recall plus extralist intrusions. A second was recall of words on a list, regardless of whether each word was reported during recall of List 1 or 2, but excluding repetitions of the word. A third measure was the total number of words produced at recall, including both extralist intrusions and repetitions. None of these revealed reliable MDR, but one measure that did show MDR was a recall composite that used the correct list recall during the final recall period, plus intrusions of those words *on the other list*, plus extralist intrusions. The significant results for this recall composite measure are shown in Table 1.

In summary, although a mood-dependent effect was found with one of many possible dependent measures, there was no evidence for a mood-dependent effect on retrieval when the simplest, most direct, measures were used.

Influences of Happy and Sad Moods on Learning

To examine the influence of mood on learning, a mood of learning (happy vs. sad) × learning time (first vs. second) × word content (pleasant vs. unpleasant) ANOVA was run. Neither happy nor sad moods showed any advantage in overall learning, and pleasant words were not recalled overall more than unpleasant words. However, mood and word content interacted strongly [$F(1,192) = p < .0001$]. Happy moods led to better learning of pleasant words, and sad moods to better learning of unpleasant words.

Influences of Happy and Sad Moods on Recall

To examine the influence of mood on recall, a recall mood (happy vs. sad) × order (first vs. second learning) × learning mood (happy vs. sad) × word content (pleasant vs. unpleasant) ANOVA was conducted for the 32 subjects who had two different moods in learning (ABA and ABB conditions). The dependent measure was the first measure (percent retention) discussed above. No main effect was significant. Two interactions reached significance, one reflecting poorer recall for the first list, the second (a three-way interaction) being uninterpretable.

DISCUSSION

To summarize, first, the mood manipulations were effective, and the moods intended as similar were indeed judged as more similar than those intended to be dissimilar. Depending on the recall measure used, there was either no evidence for MDR or slight evidence. Statistical tests that had reached significance in the earlier study failed to do so here. Furthermore, a strong word-pleasantness × learning-mood interaction was found.

These results run counter to Gilligan's earlier findings (reported in Experiment 3 of Bower et al., 1978). First, the MDR found earlier was not replicated in the present experiment. Second, a mood-congruity effect, whereby mood-congruent words were learned best, had not been found in the earlier experiment but was found in the present work. A reanalysis of Gilligan's earlier data, however, showed trends toward mood congruity.[2] Recall that about half the words used in the original experiment were replaced in our study to enhance the difference between pleasant versus unpleasant words in the lists, and this could enhance the mood-congruity effect. Furthermore, our subjects were at recall ceiling for only 8% of the learning trials, compared with 21% in Gilligan's study, perhaps due to the one fewer study trial here. Interestingly, the congruity effect occurred here despite our not expecting it, so it is not likely to have been due to experimenter demand. Better controls, however, such as tape-recorded lists and balanced presentations, would increase our own confidence in this finding.

Conflicting Evidence for Mood-Dependent Retrieval

The absence of a mood dependent retrieval effect in the present experiment is more difficult to understand in light of Gilligan's (as reported in Bower et al., 1978) earlier results. Using the same dependent measure as the earlier experiment, we found no evidence for mood-dependent retrieval. Only by using a more complicated (and suspect) measure could we find a significant MDR effect. In retrospect, the two experiments differed in several respects, of which any one might possibly account for the different results. First, the change in the word lists might conceivably cause the MDR effect to disappear. If so, MDR is surely an effect of no general importance. Second, the distractor task in the original experiment had been to read and study a book chapter. That interpolated task was replaced in the present experiment by a series of letter-cancellation tasks. This replacement might have resulted in less verbal interference with the lists learned in the present experiment, with a corresponding enhancement of a list-recency effect, and that could mask an MDR effect.

Finally, the differing results might be attributed to the hypnotic skills of the subjects and their relationship to the experimenter-hypnotist. In the earlier successful experiment, half the subjects came from an introductory psychology class, selected for a high score on a group-administered hypnosis scale, and half were mental health professionals in a long-term hypnosis workshop led by Stephen Gilligan. By contrast, all our subjects were introductory students, and this was their first ex-

perience with the experimenter-hypnotist. Possibly, Gilligan's highly skilled subjects, with their special rapport, were better able to show MDR than were our naive subjects. The skilled subjects might have shown MDR due either to more intense mood involvement or to greater compliance with apparent experimenter demand for MDR. If this hypothesis is correct, it seriously restricts MDR to a special population of people and social situations and undermines the basis for general interest in it. But this "hypnotic skill" hypothesis does not explain the other positive demonstrations of MDR, such as those by B. Thompson (cited in Bower, 1981), by Bartlett et al. (1982), and by Schare, Lisman, and Spear (1984). Accordingly, we are at a loss to explain the discrepancies among all these findings.

We conclude that mood-dependent retrieval is an evanescent will-o'-the-wisp, and not the robust outcome suggested by earlier reports. Moreover, the failure to find an MDR effect impacts negatively upon many theories that expect it. It not only contradicts Bower's (1981) specific theory of mood as an active retrieval cue; the failure impacts more generally upon *any* theory that supposes that internal states act as contexts that can become associated with memories of coincident events and can cue retrieval of them (e.g., Clark, Milberg, & Ross, 1983; Estes, 1958; Hull, 1943; Tulving & Thomson, 1973).

This spotty record of successes and failures in producing the MDR effect is assuredly an unhappy and thoroughly regrettable state of affairs. But the responsibility of investigators is not to judge their findings but simply to report them, even if at times those findings appear conflicting, contradictory, and confusing. Perhaps conflicting results are Nature's way of telling us that our questions about mood-dependent memory are formulated too simplistically.

REFERENCES

BARTLETT, J. C., BURLESON, G., & SANTROCK, J. W. (1982). Emotional mood and memory in children. *Journal of Experimental Child Psychology*, **34**, 59-76.

BARTLETT, J. C., & SANTROCK, J. W. (1979). Affect-dependent episodic memory in young children. *Child Development*, **50**, 513-518.

BOWER, G. H. (1981). Mood and memory. *American Psychologist*, **36**, 129-148.

BOWER, G. H., MONTEIRO, K. P., & GILLIGAN, S. G. (1978). Emotional mood as a context for learning and recall. *Journal of Verbal Learning and Verbal Behavior*, **17**, 573-585.

CLARK, M. S., MILBERG, S., & ROSS, J.(1983). Arousal cues material stored in memory with a similar level of arousal: Implications for understanding the effects of mood on memory. *Journal of Verbal Learning and Verbal Behavior*, **22**, 633-649.

ELLIS, E. A. (1983). *Emotional mood as a context for state-dependent retention: Some limitations of the phenomenon.* Unpublished senior honors thesis, Department of Psychology, University of Toronto.

ESTES, W. K. (1958). Stimulus-response theory of drive. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation* (Vol. 6). Lincoln: University of Nebraska Press.

GOERSS, J. C., & MILLER, M. E. (1983, May). *Memory and mood: State-dependent retention and induced affect.* Paper presented at the meeting of the Midwestern Psychological Association, Chicago, IL.

HULL, C. L. (1943). *Principles of behavior.* New York: Appleton-Century-Crofts.

ISEN, A. M., SHALKER, T., CLARK, M., & KARP, L. (1978). Positive affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of Personality and Social Psychology*, **36**, 1-12.

LEIGHT, K. A., & ELLIS, H. C. (1981). Emotional mood states, strategies, and state-dependency in memory. *Journal of Verbal Learning and Verbal Behavior*, **20**, 251-266.

MACHT, M. L., SPEAR, N. E., & LEVIS, D. J. (1977). State-dependent retention in humans induced by alterations in affective state. *Bulletin of the Psychonomic Society*, **10**, 415-418.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery, and meaningfulness values of 925 nouns. *Journal of Experimental Psychology Monographs*, **76**(1, Pt. 2).

SCHARE, M. L., LISMAN, S. A., & SPEAR, N. E. (1984). The effects of mood variation on state-dependent retention. *Cognitive Therapy and Research*, **8**, 387-407.

SHOR, R. E., & ORNE, E. C. (1962). *Harvard Group Scale of Hypnotic Susceptibility, Form A.* Palo Alto, CA: Consulting Psychologists Press.

TULVING, E., & THOMSON, D. N. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, **80**, 352-373.

WEITZENHOFFER, A. M., & HILGARD, E. R. (1962). *Stanford Hypnotic Susceptibility Scale, Form C.* Palo Alto, CA: Consulting Psychologists Press.

NOTES

1. This third experiment was Stephen Gilligan's first-year graduate project supervised by Bower. Such projects have the approximate scope of a master's thesis.

2. The possibility of coding errors in Gilligan's original data protocols could not be checked, since those protocol sheets were discarded when he moved from Stanford in 1982. We retained only records showing each subject's summary recall total for pleasant and unpleasant words for each learning or retention trial. Those summaries agree with the means reported in the Bower et al. (1978) paper.