

The bizarreness effect in a multitrial intentional learning task

KEITH A. WOLLEN and STEVEN D. COX
Washington State University, Pullman, Washington 99164

There is a widespread popular belief that bizarre imagery facilitates learning more than nonbizarre imagery. However, most research has failed to support this view. Of the few experimenters who have reported a bizarreness effect, most have used designs involving incidental learning and only a single trial. The goal of the present experiment was to use the same materials as those who reported bizarreness to be effective, but to use them in an intentional task involving several trials. Another variable added was the presence or absence of cues during recall. In general, this intentional task produced the same results as previous incidental learning tasks. For free recall, bizarre materials produced significantly greater recall than nonbizarre ones, but this difference disappeared by Trial 3. For cued recall, in contrast, nonbizarre materials produced consistently greater recall.

Memory experts have long advocated the use of bizarre imagery to facilitate the formation of associations (e.g., Cermak, 1975, p. 87; Lorayne, 1957, p. 40). According to such people, it is easier to associate, say, "bird" and "canoe" by forming a bizarre image (perhaps a bird paddling a canoe) than by forming a nonbizarre image (such as a bird perched on the edge of a canoe). Although belief in the effectiveness of bizarreness seems pervasive among memory experts, most research has shown that bizarre images are no more effective than nonbizarre ones (e.g., Nappe & Wollen, 1973; Senter & Hoffman, 1976; Wollen, Weber, & Lowry, 1972). Contrasting with these negative results, a few recent experiments seem to provide support for the effectiveness of bizarreness (e.g., Merry, 1980; Merry & Graham, 1978). The generality of Merry's data is limited, since Wollen and Cox (1981), who used Merry's materials and procedure, showed that the superiority of bizarreness was obtained only in a free recall situation; the reverse was true for cued recall.

All of the preceding experiments used only a single trial, and most used an incidental learning task. Since memory experts advocate using bizarre imagery for active learning, it seems more reasonable to use an intentional learning task. In addition, it may well be that a single trial is an artificial learning situation, since most learning probably involves multiple trials. In fact, as subjects become more experienced at forming bizarre images, one might expect to lose the superiority of nonbizarre materials over bizarre ones when cued recall is used. Accordingly, the goal of the present research was to compare free and cued recall using multiple

Appreciation is expressed to Michael R. Westbrook, who ran the subjects for this experiment. Requests for reprints should be sent to Keith A. Wollen, Department of Psychology, Washington State University, Pullman, Washington 99164.

trials and an intentional learning task. To maximize the chances of obtaining the bizarreness effect, we used the same procedure and materials that Merry found effective (Merry & Graham, 1978). In general, the procedure involved presenting bizarre or nonbizarre sentences and asking subjects to form images.

METHOD

Materials and Subjects

The materials were those used by Merry and Graham (1978). The materials consisted of 18 sentences, 6 of which were bizarre (e.g., "The man pecked a worm"), 6 nonbizarre (e.g., "The horse ate the hay"), and 6 abstract (e.g., "The others knew nothing"). The complete list of sentences is available elsewhere (Wollen & Cox, 1981, Table 1).

The subjects were 40 females and 19 males whose participation partially fulfilled a requirement of an introductory psychology course. The subjects were randomly assigned to 1 of 12 experimental conditions and were run individually. The complete design called for one more subject in five of the 12 conditions, for a total of five more subjects, but the available subjects and the school year ran out before the design could be completed. Since the design was nearly complete, and since the experimenter was leaving school, it was decided to analyze the present data rather than wait until a new year and use a different experimenter for only five subjects.

Design and Procedure

The design was mixed, with two within-subjects variables, trials and type of sentence; the major between-subjects variable was presence or absence of cuing during recall. In addition, several variables were counterbalanced; these will be described as they are introduced in the following discussion.

For Trial 1, the 18 sentences were randomly arranged with the restriction that each block of 6 sentences would contain 2 that were bizarre, 2 that were nonbizarre, and 2 that were abstract. A different random order was used for each subsequent trial. One counterbalanced variable was which one of two input orders was used. The sentences were presented on a 12-in. monitor using a TRS-80 computer. Each sentence was presented for 8 sec, and subjects were asked to form a mental image of the events or relationships described in the sentences.

A distractor task was given after input of all 18 sentences. The distractor consisted of a series of sentences that appeared one at a time on the monitor. The subject's task was to count the number of times the letter "e" appeared in each sentence and then say that number aloud. The experimenter recorded the number and presented the next sentence. The subjects were told that the object was to see what effect the formation of images had on the speed and accuracy of letter counting and on the learning of the sentences. Thus, an intentional learning task was used instead of the incidental tasks frequently employed in previous research (Merry, 1980; Merry & Graham, 1978; Wollen & Cox, 1981).

After 1 min of the distractor activity, subjects in the free recall conditions saw the word "RECALL" on the screen. That was their cue to take a response sheet from a nearby pile and write as many of the sentences as they could. Subjects were asked to write the sentences as close to the original wording as possible, but to write whatever they could, even if only one or two words.

Subjects in the cued recall conditions were treated identically, except they had the first two words of each sentence provided as cues (e.g., "The man," "The horse," etc.) The cues were presented on the screen in nine rows of two cues each. A different random sequence was used on each subsequent trial. There were two different sets of sequences that constituted another balanced variable. The recall period was unpaced and ended whenever the subject put his or her response sheet through a slot in a 4 x 4 ft panel separating the experimenter from the subject. This procedure continued until the subject reached a criterion of one perfect trial or until the hour had elapsed, whichever came first. In both the cued and the free recall conditions, the subjects were free to write the sentences in whatever order they wished.

RESULTS AND DISCUSSION

The data were first examined in a way that was most similar to the scoring used by Merry (1980), which is known to produce the bizarreness effect. Each sentence contained a first noun (N1), a verb (V), and a second noun (N2). Merry counted as correct both N1 and N2, regardless of whether they were correctly paired. We scored N2 recalls as correct irrespective of whether they were paired with the correct N1 responses. The N1 responses were not considered because the experimental design includes both cued and free recall and the N1 items were provided as cues for the cued conditions. Hence, to make free and cued recall comparable, we could only consider N2 responses. The number of correct N2 recalls was determined for each subject and for each trial. The means of these values for the various conditions are shown in Figure 1.

Most other experimenters used only one trial, and so the data most relevant to those studies are those from Trial 1. The Trial 1 data were analyzed by a mixed-design analysis of variance, with the between-subjects variable being type of recall (free or cued) and the within-subjects variable being type of sentence (bizarre or nonbizarre).

There was a significant interaction between type of recall and type of sentence [$F(2,114) = 8.96, p < .001$]. Inspection of Figure 1 reveals that the interaction arises primarily because bizarre sentences increased free recall relative to nonbizarre sentences (BF > NF) but decreased

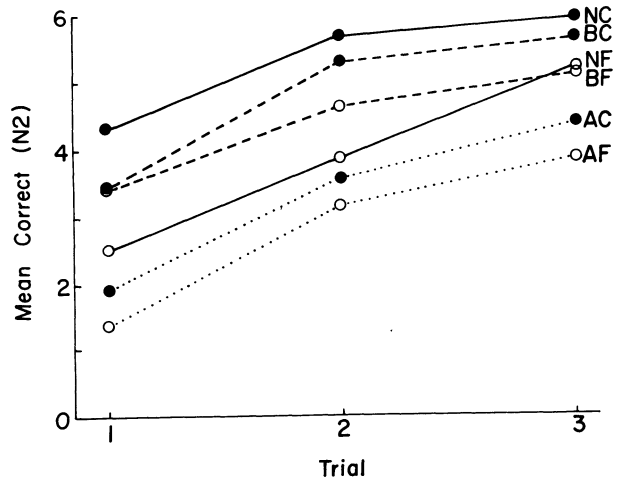


Figure 1. Mean number of correct N2 responses as functions of type of sentence (B, N, or A), type of recall (C or F), and trial.

cued recall (BC < NC). The respective values of t were 2.98 and 2.80, which are significant at the .01 level ($df = 29$ and 28).

The superiority of BF over NF replicates recent findings that bizarreness is effective in a free recall paradigm (Merry, 1980; Merry & Graham, 1978). Since we used an intentional learning task, the superiority of BF to NF also shows that the bizarreness effect in free recall is not dependent upon the use of the incidental task used by Merry. The finding that bizarreness produced lower recall when cues were provided (BC < NC) is in line with previous results from an incidental task (Wollen & Cox, 1981).

The other point of interest is what happens to the bizarreness effect over trials. Figure 1 shows data for only three trials because that is the number most people completed. Four subjects reached the criterion of one perfect trial in only two trials; these subjects were treated as though they would have had perfect scores on Trial 3 as well. A total of 22 subjects did not reach criterion in three trials but were not given a fourth trial because the experimental hour had elapsed. Consequently, data can be presented only for three trials.

As can be seen in Figure 1, the bizarreness effects do not seem to change much over trials. In free recall conditions, BF starts out above NF but ends up slightly lower by Trial 3. In cued recall, NC remains superior throughout. The maximum number correct was six, and several means approached that value. This resulted in arrays of scores that were not amenable to analysis of variance because normality and homogeneity assumptions could not be satisfied. However, it does not appear that the pattern of results changes as a result of multiple trials, at least within the range employed herein.

The results were also examined in terms of the number of times that N2 was correctly paired with N1 and in terms of the number of times that the entire

sentence was recalled correctly. These analyses produced slightly lower overall levels of correct responses, but the resulting curves looked nearly identical to those shown in Figure 1.

Since multiple trials were used, it was possible to examine the trials taken to reach the criterion of all six bizarre sentences correct and all six nonbizarre sentences correct. It would be expected that the trials to criterion and mean number recalled would yield similar results, and such was the case. For purposes of this analysis, only the bizarre and nonbizarre sentences were included; too few people reached criterion for the abstract sentences to be considered. The number of trials to criterion was determined separately for bizarre and nonbizarre sentences. If subjects did not reach criterion in the trials given, they were assigned a score equal to what their score would have been had they reached criterion on the next trial. Thus, if a subject recalled four of six bizarre sentences on Trial 4 and had no Trial 5, he or she was given a score of 4. It was only necessary to do this 19 times out of 116 that were possible, thereby indicating that the criterion was usually reached for both bizarre and nonbizarre sentences. One subject was much slower than any other and finished only two trials during the entire session; this subject did not reach either criterion, and these data were not considered in the trials-to-criterion analysis.

The data were analyzed by a mixed analysis of variance; the within-subjects factor was type of sentence (bizarre or nonbizarre) and the between-subjects factor was type of recall (free or cued). There was a significant interaction between type of sentence and type of recall [$F(1,56) = 11.72, p < .01$], and so the simple effects were examined. For free recall, bizarre sentences required significantly fewer trials to criterion than nonbizarre

sentences [2.20 vs. 2.59; $t(29) = 2.03, p \cong .05$]. For cued recall, the opposite was obtained; bizarre sentences required significantly more trials to criterion [1.72 vs. 1.17; $t(29) = 2.86, p < .01$]. Hence, the trials-to-criterion data show the same pattern as the mean number correct.

In summary, bizarre sentences produce superior performance only in free recall, and that superiority breaks down by the third trial. In cued recall, bizarre sentences produce consistently poorer performance. Consequently, the memory experts' advocacy of bizarreness seems to be reasonable only when there will be no retrieval cues at the time of recall.

REFERENCES

- CERMACK, L. S. *Improving your memory*. New York: McGraw-Hill, 1975.
- LOBAYNE, H. *How to develop a super-power memory*. New York: Fell, 1957.
- MERRY, R. Image bizarreness in incidental learning. *Psychological Reports*, 1980, *46*, 427-430.
- MERRY, R., & GRAHAM, N. C. Imagery bizarreness in children's recall of sentences. *British Journal of Psychology*, 1978, *69*, 315-321.
- NAPPE, G. W., & WOLLEN, K. A. Effects of instructions to form common and bizarre mental images on retention. *Journal of Experimental Psychology*, 1973, *100*, 6-8.
- SENER, R. J., & HOFFMAN, R. R. Bizarreness as a nonessential variable in mnemonic imagery: A confirmation. *Bulletin of the Psychonomic Society*, 1976, *7*, 163-164.
- WOLLEN, K. A., & COX, S. D. Sentence cuing and the effectiveness of bizarre imagery. *Journal of Experimental Psychology: Human Learning and Memory*, 1981, *7*, 386-392.
- WOLLEN, K. A., WEBER, A., & LOWRY, D. H. Bizarreness versus interaction of mental images as determinants of learning. *Cognitive Psychology*, 1972, *3*, 518-523.

(Received for publication October 1, 1981.)