Sentence complexity eliminates the mnemonic advantage of bizarre imagery

MARK A. McDANIEL
Purdue University, West Lafayette, Indiana

and

GILLES O. EINSTEIN
Furman University, Greenville, South Carolina

Research has consistently shown that *bizarre* sentences are better remembered than *common* sentences when bizarreness is manipulated within-subjects, memory is assessed with free recall, and the processing instructions emphasize the use of imagery. Even though previous research conforms remarkably to this pattern, a recent exception to this pattern (Kroll, Schepeler, & Angin, 1986, Experiment 1) suggests that there are additional factors that control bizarre imagery effects. Several methodological factors were experimentally examined to try to identify the reason for the discrepancy between the Kroll et al. findings and the rest of the literature. The results indicated that the complexity of the stimulus materials can eliminate the bizarreness effect. Several explanations for this finding are discussed.

It is well known that *bizarre* imagery does not invariably provide mnemonic benefits, relative to imagery that is not bizarre (labeled *common imagery* in the present paper). Accordingly, recent research efforts have focused on delineating the conditions under which bizarre imagery will and will not have positive effects on memory performance (e.g., Einstein, McDaniel, & Lackey, in press; Kroll & Tu, 1988; McDaniel & Einstein, 1986). One factor that has been consistently associated with whether or not bizarre imagery effects are observed is the nature of the experimental manipulation. Between-subjects manipulations of imagery type (bizarre or common) generally do not produce bizarre imagery effects (although see Einstein et al., in press), whereas within-subjects manipulations of imagery type produce better free recall for bizarre relative to common images. The latter finding is so robust that it has been reported in every study using within-subjects manipulations (in which imagery processing is required) except one (see Einstein & McDaniel, 1987, for a review of nine studies in which bizarreness has been manipulated within-subjects and memory was tested with free recall). Specifically, Kroll, Schepeler, and Angin (1986) recently failed to find any mnemonic advantage of bizarre imagery despite using a within-subjects manipulation and measuring free recall (Experiment 1). We conducted the present research to further our understanding of the limits of the bizarreness effect. Our approach involved exploring why Kroll et al. (1986) did not obtain benefits of bizarre imagery in free recall using a within-subjects design.

In trying to reconcile the Kroll et al. (1986) findings with the numerous demonstrations of positive effects of bizarre imagery, Einstein and McDaniel (1987) identified several potentially important methodological differences: list length, orienting tasks, and stimulus materials. Einstein and McDaniel argued that the list length variable was probably not critical (on the basis of a recently completed study by McDaniel, Anderson, Einstein, & O'Halloran, in press), but left open the possibility that the orienting tasks (examined in Experiment 1) and/or the stimulus materials (examined in Experiment 2) were critical factors in obtaining the effect.

EXPERIMENT 1

One possibility for the absence of bizarre imagery effects in Kroll et al. (in free recall) concerns the orienting task under which subjects encode the to-be-imaged material. One novel feature of the Kroll et al. study (1986, Experiment 1) was that subjects were instructed to rate each of their images (of the events described by sentences) for bizarreness, degree of interaction, and vividness. Bizarre imagery studies in which subjects rate images typically have them rate the images on one dimension only. There are at least two reasons why the Kroll et al. multiple rating task might have eliminated the mnemonic advantage for bizarre imagery. One reason is that rating images on three dimensions may increase the distinctiveness of the encodings (Battig & Einstein, 1977; Klein & Saltz, 1976). Assuming that bizarre imagery is beneficial for recall because it produces a distinctive encoding (Einstein

& McDaniel, 1987; Einstein et al., in press; McDaniel & Einstein, 1986), then performing a multiple rating task (which itself enhances the distinctiveness of the encoding) could attenuate the advantage of image bizarreness.

Another reason is that the multiple rating task involved the rating of bizarreness, a task that McDaniel and Einstein (1986) speculated may reduce imaginal processing (a bizarreness effect does not seem to obtain with semantic orienting tasks; see McDaniel & Einstein, 1986; McDaniel et al., in press). Some support for this hypothesis has been recently reported by Kroll and Tu (1988), who failed to find a significant advantage for bizarre imagery with a bizarreness rating task, but did find a significant advantage with a vividness rating task.

In Experiment 1, we directly compared the effects of bizarre imagery (relative to common imagery) when the orienting task required a vividness rating only, as opposed to when the task required a vividness rating, a bizarreness rating, and a rating of degree of interaction.

Method

Subjects and Design. Thirty-two students from an introductory psychology course at Furman University participated for course credit, with 16 subjects assigned to each of the two orienting task conditions (one rating vs. three ratings). The imagery-type variable (bizarre vs. common) was manipulated within-subjects.

Procedure. The sentences were those used in McDaniel and Einstein (1986, Experiments 1–4; see their Appendix). Target noun triplets were embedded in either bizarre or common sentence frames, with the target nouns presented in capital letters. Although each subject saw each noun triplet in only a bizarre sentence or a common sentence, across subjects each noun triplet was embedded equally often in a bizarre and a common sentence frame. The sentences were presented on a CRT for 10 sec each and in a random order (with the restriction that no more than two sentences of one type appeared adjacently). All the subjects were instructed to form a visual image of the event depicted in the sentence during the 10 sec. After each 10-sec presentation period, the rating scale(s) appeared on the screen, and the subjects were allowed as much time as was needed to complete their rating(s).

The subjects in the single-rating condition rated the vividness of their images on a 5-point scale (with 1 indicating a clear, vivid image, and 5 indicating a poor, unclear image). The subjects in the multiple-rating condition first rated vividness, then they rated the degree of bizarreness of each image (with 1 indicating common, and 5 indicating bizarre), and finally they rated the degree of interaction among the target (capitalized) words (with 1 indicating no interaction, and 5 indicating high interaction). After processing the sentence set, the subjects solved math problems for 1 min, and then they were given a free recall test in which they were asked to write down as many of the capitalized nouns as they could. The subjects were not forewarned about this memory test.

Results and Discussion

All statistical tests were performed at the .05 level. A two-factor mixed analysis of variance (with rating condition as the between-subjects variable and imagery type as the within-subjects variable) on the vividness ratings showed that bizarre images were rated as less vivid than common images [2.6 vs. 1.7; $F(1,30) = 25.83$, $MSe = .52$]. There were no other effects in this analysis ($Fs < 1$). The subjects that performed three ratings rated bizarre sentences as more bizarre than common sentences [3.9 vs. 2.3; $F(1,15) = 21.07$, $MSe = 1.03$], but rated the sentences equivalently in terms of interactiveness [3.1 vs.

3.6 for bizarre and common, respectively; $F(1,15) = 1.82$, $MSe = .87$].

A two-factor mixed analysis of variance (like that performed for vividness ratings) on the number of nouns recalled (see Table 1 for mean proportion recalled) revealed that recall was significantly better for bizarre than common imagery [$F(1,30) = 7.09$, $MSe = 9.60$]. This effect held for both rating conditions ($F < 1$ for the interaction), and recall was equivalent across both rating conditions ($F < 1$).

Thus, it appears unlikely that the Kroll et al. (1986) rating task was responsible for their failure to obtain a bizarre imagery effect in free recall. This is not to say that the particular orienting task is unimportant in obtaining bizarre imagery effects; sentences that depict bizarre events are not recalled better (than sentences that depict common events) when encoded with orienting tasks that do not emphasize imagery (McDaniel & Einstein, 1986, Experiment 2; McDaniel et al., in press).

EXPERIMENT 2

As noted in our introduction, another feature of the Kroll et al. (1986) study that might have contributed to the absence of bizarre imagery effects was the particular stimulus materials used. To exercise control over the types of images that subjects form, recent studies (including the nine reviewed in Einstein & McDaniel, 1987) typically have presented subjects with target noun pairs or noun triplets embedded in sentences. Kroll et al. also presented noun pairs within sentence frames, but the sentences describing the to-be-imaged events (except for one set of control sentences) tended to contain many modifiers and details (e.g., "A large, round BOULDER bends and groans under the weight of the old, tired FROG" was a bizarre sentence, and "A large, round BOULDER rolls relentlessly over the unfortunate, old FROG" was a corresponding common sentence; see the Appendix in Kroll et al. for the complete list). This degree of complexity is a departure from the kinds of simple, unembellished sentences other researchers have used (e.g., "The MAID licked AMMONIA off the TABLE" from McDaniel & Einstein, 1986; "The HEN smoked a CIGAR" from Merry, 1980). Sentence complexity could well influence the bizarre imagery effect by making common sentence encodings as distinctive as bizarre sentence encodings. To test the idea that the complexity of the sentence frame might influence the degree to which bizarre imagery effects would be manifested, in Experiment 2 we manipu-

Table 1
Mean Proportion of Target Nouns Recalled in Experiment 1

	Type of Imagery	
Orienting Task	Bizarre	Common
Vividness Rating only	.58	.47
Vividness Rating, Bizarreness Rating, and Interactiveness Rating	.61	.49

lated sentence complexity for both the Kroll et al. materials and the McDaniel and Einstein materials.

Method

Subjects and Design. Forty-eight undergraduate students at Notre Dame participated for extra credit toward their introductory psychology course. The design was a three-factor mixed factorial, with sentence complexity (simple vs. complex) and number of target nouns (noun pairs vs. noun triplets; the Kroll et al. sentences contained noun-pair targets and the McDaniel & Einstein sentences contained noun-triplet targets) as the between-subjects factors and sentence bizarreness (bizarre vs. common) as the within-subjects factor. Twelve subjects were randomly assigned to each of the four groups defined by the combination of the between-subjects factors.

Materials. A stimulus list of 12 sentences was constructed for each of the four experimental groups. The simple sentences containing noun triplets were taken directly from McDaniel and Einstein (1986; see their Appendix); they averaged 8 words per sentence. The complex sentences containing noun triplets were constructed by adding adjectives to each of the simple sentences (e.g., "The BANKER dropped the NEWSPAPER in the PUDDLE" was changed to "The tired, crabby BANKER dropped the torn, crumpled NEWSPAPER into the muddy PUDDLE"); they averaged 13 words per sentence. The complex sentences containing noun pairs were adopted directly from Kroll et al. (1986; a subset of the plausible and bizarre sentences listed in their Appendix was used); they averaged 11 words per sentence. Also, the simple common sentences containing noun pairs were adopted directly from Kroll et al. (the plausible-short sentences in their Appendix). Kroll et al. did not use simple bizarre sentences, so we constructed simple bizarre sentences by removing the modifiers from the corresponding complex bizarre sentences. The simple sentences containing noun pairs averaged 6 words per sentence.

Each stimulus list contained 6 bizarre and 6 common sentences randomly ordered, with a different random order prepared for each subject. All lists were counterbalanced so that across subjects each sentence frame was presented in its bizarre form and in its common form equally often.

Procedure. The procedure was the same as that for the vividness rating condition in Experiment 1, except that sentence presentation was experimenter-paced. The presentation rate was based on the average number of words for each sentence set. Following McDaniel and Einstein (1986, Experiment 4), the simple sentences containing noun triplets were presented for 7 sec. This was used as the base rate for determining the presentation rate for the other sentences: 13 sec for the complex sentences containing noun triplets, 5 sec for the simple sentences containing noun pairs, and 11 sec for the complex sentences containing noun pairs. Following the presentation of each sentence, the subjects were given 4 sec to rate the vividness of their image, after which the next sentence appeared. After imaging and rating the list of sentences, the subjects solved embedded-figure problems for 5 min. Then the subjects were given 4 min to recall as many of the target (capitalized) words as they could.

Results

The rejection level for all tests was set at .05. A three-factor analysis of variance of the vividness ratings indicated that bizarre images were rated as significantly less vivid than common images [2.7 vs. 1.9; $F(1, 44) = 36.88$, $MSe = .36$]. There were no other effects (all $Fs < 1$).

The proportion of target nouns recalled (see Table 2 for means) was analyzed with a three-factor mixed ANOVA (sentence complexity and number of target nouns as between-subjects factors and sentence bizarreness as the within-subjects factors). There was no main effect of sentence bizarreness ($F < 1$), but sentence bizarreness significantly interacted with sentence complexity [$F(1, 44)$ $= 7.01$, $MSe = .03$]. Examination of Table 2 reveals that

Table 2
Mean Proportion of Target Nouns Recalled in Experiment 2

Sentence Complexity	Type of Imagery		
	Target Type	Bizarre	Common
Simple	Noun pairs	.53	.46
	Noun triplets*	.49	.37
Complex	Noun pairs†	.40	.51
	Noun triplets	.22	.30

*McDaniel & Einstein (1986) materials;   †Kroll et al. (1986) materials.

this interaction was due to nouns from bizarre sentences being better recalled than nouns from common sentences when the sentence frames were simple, with this pattern reversing itself when the sentence frames were complex. This pattern obtained for both noun-triplet targets (McDaniel & Einstein, 1986, materials) and noun-pair targets (Kroll et al., 1986, materials), with $F < 1$ for the three-way interaction. Two other significant effects emerged from this analysis. Simple sentences produced better memory than complex sentences [$F(1, 44) = 6.46$, $MSe = .04$], and a higher proportion of nouns were remembered from the noun pairs than the noun triplets [$F(1, 44) = 9.81$, $MSe = .04$].

Discussion

The results were clear cut. With the Kroll et al. (1986) complex sentences, there was no mnemonic advantage in free recall for bizarre imagery, as reported by Kroll et al. (Experiment 1). In fact, for this sentence set, common sentences were recalled better than bizarre sentences. This same pattern has recently also been found in an unpublished study (Hirshman & Whelley, 1988, Experiment 1), in which simple common sentences were used as a comparison for the bizarre sentences. When the Kroll et al. sentence frames were simplified, however, bizarre imagery benefited free recall. Similarly, with the McDaniel and Einstein (1986) simple sentences, bizarre imagery benefited free recall, as also reported by McDaniel and Einstein and replicated by Hirshman and Whelley. When these simple sentence frames were made more complex, however, the mnemonic advantage of the bizarre sentences disappeared. Thus, it is likely that the complexity of Kroll et al.'s sentences was responsible, at least in part, for their failure to obtain benefits of bizarre imagery.

At this time it is not clear why complex sentence frames like those used by Kroll et al. (1986) eliminate the advantage of bizarre imagery. One possibility is that the increased complexity reduces the imageability of the event depicted. Consistent with this hypothesis, Kroll et al. found that the complex sentences produced significantly lower imagery vividness ratings than did short, plausible sentences. Thus, instead of increasing image content as intended (cf. Kroll et al., 1986), the complex sentences may have decreased visual image content for some as yet undetermined reason (it is interesting that the extra modifiers inserted in the Kroll et al. sentences do not always seem to be easily instantiated in a visual image—e.g., "rolls relentlessly," "unfortunate," "bends and groans," "tired" from the sentence examples above). To the extent that the mnemonic benefit of bizarre sentences seems to be dependent on imagery encoding (McDaniel & Einstein, 1986; McDaniel et al., in press), such a decrease in imagery processing would be expected to eliminate the mnemonic advantage of sentences depicting bizarre events. The absence of vividness-rating differences across sentence complexity in the present study does not support the above interpretation, however.

Another possibility rests on the observation that the additional modifiers in the complex sentences produce more elaboration of the target nouns. Assuming that elaboration results in a more distinctive encoding (Stein, Littlefield, Bransford, & Persampieri, 1984), then it may be that complex sentences produce encodings of common sentences that

are as distinctive as those of bizarre sentences. This interpretation is similar to Einstein and McDaniel's (1987) observation that the degree of difference between bizarre and common sentences may be less pronounced in complex sentences such as those used by Kroll et al. (1986). The finding that complex sentences did not produce better recall than simple sentences is not necessarily inconsistent with the foregoing idea. Although elaboration may increase distinctiveness, memory performance is not enhanced unless the elaboration makes the relationship between the target concepts less arbitrary (Stein et al., 1984). Informal examination of our target sentences (e.g., see Kroll et al., 1986, Appendix) indicates that the additional modifiers in the complex sentences do little to enhance (and in some cases, may obscure) the precision of the relationships.

At any rate, the important point from the present research is that bizarre imagery effects in free recall can be reliably obtained in within-subjects designs, provided that the stimulus sentences are not too complex. Further research is necessary to determine whether sentence complexity impedes image formation or leads to encodings of common targets that are well elaborated and perhaps as a consequence more distinctive. Such information would be useful for evaluating theoretical interpretations of bizarre imagery effects.

REFERENCES

BATTIG, W. F., & EINSTEIN, G. O. (1977). Evidence that broader processing facilitates delayed retention. *Bulletin of the Psychonomic Society*, **10**, 28-30.

EINSTEIN, G. O., & McDANIEL, M. A. (1987). Distinctiveness and the mnemonic benefits of bizarre imagery. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes: Theories, individual differences, and applications* (pp. 78-102). New York: Springer-Verlag.

EINSTEIN, G. O., McDANIEL, M. A., & LACKEY, S. (in press). Bizarre imagery, interference, and distinctiveness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.

HIRSHMAN, E., & WHELLEY, M. M. (1988). *An investigation of paradoxical memory effects.* Unpublished manuscript, New York University, New York, NY.

KLEIN, K., & SALTZ, E. (1976). Specifying the mechanisms in a levels-of-processing approach to memory. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 671-679.

KROLL, N. E. A., SCHEPELER, E. M., & ANGIN, K. T. (1986). Bizarre imagery: The misremembered mnemonic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 42-53.

KROLL, N. E. A., & TU, S.-F. (1988). The bizarre mnemonic. *Psychological Research*, **50**, 28-37.

McDANIEL, M. A., ANDERSON, D. C., EINSTEIN, G. O., & O'HALLORAN, C. (in press). Modulation of environmental reinstatement effects through encoding strategies. *American Journal of Psychology*.

McDANIEL, M. A., & EINSTEIN, G. O. (1986). Bizarre imagery as an effective memory aid: The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 54-65.

MERRY, R. (1980). Image bizarreness in incidental learning. *Psychological Reports*, **46**, 427-430.

STEIN, B. S., LITTLEFIELD, J., BRANSFORD, J. D., & PERSAMPIERI, M. (1984). Elaboration and knowledge acquisition. *Memory & Cognition*, **12**, 522-529.