

IMAGES AS MEMORY AIDS: IS BIZARRENES HELPFUL?

Stanley C. Collyer, John Jonides, and William Bevan
Johns Hopkins University

Eighty subjects learned noun-verb-noun triplets suggesting either common, plausible scenes or bizarre, implausible scenes and under instructions to use imagery or a verbal-memorization strategy. Clearly the best free-recall performance was by the subjects who used plausible images as mediators, a superiority that was not due simply to a difference in semantic meaningfulness between the common and bizarre triplets.

The role of mediating images as aids to memory is currently enjoying a resurgence of interest, as may be seen in the recent reviews of Paivio (1969) and Bugelski (1970). A number of experiments have shown that when subjects are given instructions to use images as mediators in a paired-associates situation, recall scores are remarkably high. For example, the frequently cited study of Wallace, Turner, and Perkins (1957) reported approximately 95% correct recall with 700 paired associates, each seen only one time. More recently, Bugelski, Kidd, and Segmen (1968) also reported that, with enough time, subjects given instructions to form images outperformed controls given no such instructions, a result confirmed by Bugelski (1968) using a more difficult task. Similarly, Paivio (1968) has repeatedly shown image instructions to be superior to rote-repetition instructions or to no instructions at all. Furthermore, the assertion that subjects, when instructed to, do indeed use images instead of verbal mediators was supported indirectly by Bugelski (1968), who examined his subjects' descriptions of their mediators, and in a different fashion by Atwood (1969), who showed that visual interfering stimuli were more detrimental than auditory ones when subjects were instructed to use images, while the reverse was the case in a verbal-learning task not involving imagery. It thus would seem to be well established that images can be used effectively as mediators for learning verbal material.

Attention has now turned to specifying the characteristics of a 'good' mnemonic image—to those qualities that make it most effective as a mediator. Atwood (1969), for example, has demonstrated the importance

of *figural unity* in images. The interdependence of the various parts of the image enhances the likelihood that one part will be able to evoke the whole.

One qualitative aspect of images traditionally thought to be related to mnemonic effectiveness is their *bizarreness*, or implausibility. Advice to construct bizarre and unusual mental pictures can be found in the classical literature (see Yates, 1966), as well as in modern sources (Miller, Galanter, and Pribram, 1960). However, current attempts to confirm bizarreness as an important variable have so far produced inconclusive results (Atwood, 1969; Wood, 1967; Bugelski, 1970), probably due in part to the difficult methodological problems involved. One of these problems involves manipulating the degree of bizarreness of the images without destroying any relevance the mnemonic exercise might have for a real situation in which the subject is trying to remember a number of things. On the one hand, when the experimenter completely specifies for the subject the details of the mnemonic image, the situation is quite different from the latter's being left free to create his own. Although Briggs, Hawkins, and Crovitz (1970) did obtain high recall scores when images were specified, performance might well have been better if the subjects had been required to construct images of their own choosing. On the other hand, simply telling the subjects to construct either common or bizarre images leads to difficult description and rating problems. Furthermore, it probably also confounds figural unity with bizarreness, since the most 'common' image involving two or more initially unrelated objects would likely be a simple juxtaposition.

The experiment now reported sought to clarify the relationship between bizarreness of the mnemonic image and efficiency of free recall by following a procedural middle course. Our aim was to maintain high figural unity and, at the same time, to provide just enough guidance in the construction of the image to ensure that degree of bizarreness was manipulated even as the subject was free to construct an image mostly of his own choosing.

METHOD

—Experimental approach—The procedure involved presenting noun-verb-noun triplets with the nouns held constant across groups; the verbs differed to guide the formation of either common or bizarre images. The subject's task was to recall as many of the second nouns as possible within a given period of time. The basic approach, then, was to use a simple two-group between-subjects design. However, certain additional controls were required. While elements of the triplets might be equated for frequency of usage, associabil-

ity, or some similar criterion, the semantic meaningfulness of each triplet quite probably depends on its composition as a whole. Consequently, ease of recall of the second nouns might well vary as a function of the meaningfulness of the verbal combinations used, variations which could obscure differential recall scores due to imagery. Accordingly, it was deemed necessary to replicate the simple two-group design with another two groups given instructions to make use of a verbal-mediation strategy instead of an imagery strategy in memorizing the material. The influence of bizarreness in *image-mediated* recall is thereby indicated if the difference in performance between the two imagery groups is reliably larger than the difference for the two verbal groups.

—Subjects—In all, 140 Johns Hopkins undergraduates participated as subjects. Sixty of these were used in two preliminary exercises pertaining to stimulus preparation: the remaining 80, separated randomly into groups of 20, constituted the four cells of the 2×2 between-groups design. All were volunteers paid for their services. Each was tested individually in a 30-min session.

—Stimuli—Each of the four experimental groups received a set of 20 noun-verb-noun triplets. For two groups, pairs of nouns were connected by a verb which suggested a 'common, or plausible, relationship; for the remaining two, the same nouns were connected by a verb suggesting a 'bizarre,' or implausible, relationship.

The triplets were constructed by first selecting 44 nouns with high imagery and concreteness ratings from the Paivio, Yuille, and Madigan list (1968). On a seven-point scale, mean imagery was 6.61 and mean concreteness 6.89. Next, for each noun a common transitive verb was chosen which was *appropriate* to that noun, that is, one which naturally or frequently occurs in conjunction with the noun. Thus, for example, the verb *PULL* was paired with the stimulus noun *HORSE*, *BLAST* with *SHOTGUN*, *MEASURE* with *TABLESPOON*, and so on. Finally, 44 response nouns (mean imagery, 6.56; mean concreteness, 6.88) were selected to complete the triplets. These nouns were chosen so that they were not, in the judgment of the experimenters, highly frequent or stereotyped associates of the stimulus noun and verb.¹ At the same time, they suggested a relationship that might reasonably occur in real life.² Examples are *HORSE PULL ROCK*, *SHOTGUN BLAST BUILDING*, and *TABLESPOON MEASURE LEMONADE*.

When the 44 common triplets had been constructed, they were divided at random into two sets and the verbs interchanged between the sets, yielding two sets of bizarre triplets with 22 triplets in each set. Examples are *HORSE AWAKEN ROCK*, *SHOTGUN MEASURE BUILDING*, and *TABLESPOON BLAST LEMONADE*.

Next, the total of 88 triplets was rated for bizarreness by two groups of ten judges, each group rating 22 common and 22 bizarre triplets. This rating was necessary not only to verify that bizarreness had been manipulated in the desired fashion but also to permit the elimination of triplets for which there was no general agreement or for which subjects could not construct an image. The term 'bizarre' was defined for the judges by presenting such synonyms as 'grotesque,' 'implausible,' 'incongruous,' 'ludicrous,' 'odd,' and 'strange.' Each judge was asked to imagine a scene suggested by each triplet and to rate his image on a seven-point scale from least to most bizarre. He was cau-

tioned not to rate his difficulty in constructing an image but rather its bizarreness once constructed. As a result of these ratings, 8 triplets were discarded, and four groups of 20 remained, two composed of common and two of bizarre triplets. The mean ratings for these groups were: Common₁ = 1.70, Common₂ = 1.71; Bizarre₁ = 4.83, Bizarre₂ = 4.82.

Tests of free recall were then conducted for these four groups (with ten subjects per group), using the procedure described below for the imagery groups. As anticipated, the means were identical across lists employing different verbs [Common₁ + Bizarre₂ = Common₂ + Bizarre₁, $t(38) = .152$, *n.s.*]. As a result, Common₁ + Bizarre₂ were discarded; the main experiment proceeded with only two stimulus lists.

—Procedure—Each subject in the two *imagery groups* was told that he was in an experiment on the use of images as aids to memory. He was informed that he would receive a booklet containing a series of noun-verb-noun triplets, one triplet on each page, and that he would later be asked to recall the second noun of each triplet. His initial task was described as constructing a 'mental picture' of a scene in which the first noun did what the verb specified to the second noun (an example was given, along with a description of an image meeting this requirement). He was instructed to concentrate on each triplet until a satisfactory image was constructed, at which time he was to turn the page to the next one. The maximum time permitted for each triplet was 30 sec, with the experimenter recording the time spent on each page.

Immediately after exposure to the 20 triplets, the subject performed a 4-min distraction task, which consisted of searching for targets on a sheet of random letters. Next, a 4-min free-recall period was provided during which he wrote down all the response nouns he could remember, without regard to their order of presentation.

Each subject in the *verbal-mediation groups* was told that he was in an experiment on the use of verbal associations as aids to memory. Suggestions of verbal strategies were given (e.g., constructing a short sentence containing the three elements of the triplets, or repeating the triplets by rote). All other procedural details were identical to those for the imagery groups.

In order to verify that all the subjects had followed their respective instructions in attempting to memorize the material, each subject was interrogated after the recall test. The experimenter repeated each triplet and asked the subject to report what he did to try to memorize that triplet. If he had failed to use the designated strategy on 8 or more of the 20 items, his data were discarded. On this basis, data from 24 subjects were excluded; 17 of them were from the verbal-strategy groups, and 7 from the imagery groups. Additional subjects were run until each of the two instructional groups had 40.

RESULTS

Free recall

Figure 1 presents the mean number of response nouns correctly recalled by the several experimental groups. Inspection reveals that subjects who had received the common triplets performed distinctly better

during free recall than those presented the bizarre triplets [mean recall difference between types of triplets = 3.2 items; $F(1, 76) = 42.60$, $p < .001$]. Furthermore, subjects who had been instructed to create images from the triplets generally recalled more items than those required to use a verbal strategy [mean recall difference between types of instruction = 2.7 items; $F(1, 76) = 30.30$, $p < .001$]. Finally, the interaction between stimulus material and instructions was also significant [$F(1, 76) = 9.85$, $p < .01$].

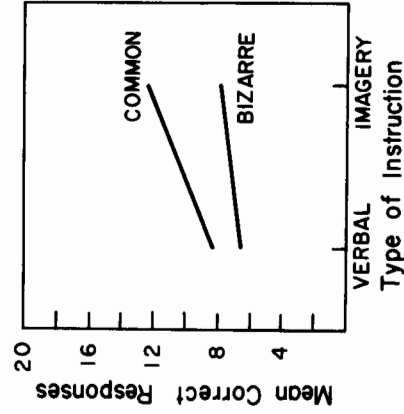


Fig. 1. Mean number of response nouns correctly recalled

The interaction quite clearly was due to the superiority of the common-imagery group. Mean separation tests bear out this conclusion. The mean differences between the group using common images and the other groups were, in order of mean amount recalled, 4.2, 4.7, and 5.9 items respectively [least significant difference ($p < .001$) = 2.36; see Kirk, 1968]. All other differences failed significance. The procedural implications of this interaction are of particular note. Had a significant interaction not obtained, it could be contended that subjects who received the common triplets displayed better free-recall performance solely because these triplets were more semantically meaningful than the bizarre triplets.

Time to construct images

An analysis of the average amount of time spent on each triplet showed no significant differences for the four experimental groups. The mean times ranged from 20.5 sec for the common-imagery group to 23.6 sec

for the bizarre-verbal group. This suggests that the superiority of common images is not simply a consequence of the greater time spent in the construction of those images.

DISCUSSION

The present data indicate that use of common or plausible triplets resulted in superior recall, with unequivocally the best performance being identified with the construction of images from this material. Thus it would appear that the frequently recommended strategy of using bizarre imagery as a mnemonic aid is unwise.

Before speculating on the reasons for this finding, let us explore the possibility that an explanation not involving imagery might easily account for our results. The likelihood that the superiority of the common-imagery group is attributable to some 'verbal' factor rather than to the content of the subjects' visual images was fairly certainly eliminated by the several procedural controls employed. First, the fact that some words are easier to remember than others was offset by using the same stimulus and response nouns in both the common and bizarre triplets. Second, while the meaningfulness of the triplet as a whole changes when the verb is changed, the inclusion of groups with identical stimulus materials, but with instructions to use a verbal-mediational strategy, makes possible the segregation of semantic effects from effects due to imagery. Indeed, the significant interaction of stimulus material and instructions makes clear the operation of nonverbal—here, specifically imaginal—factors in the enhanced recall of the common-triplet response nouns.

How should we account for the superiority of common over bizarre imagery? One might suppose, given the virtually limitless capacity of visual memory (see Wallace et al., 1957; Standing, Conezio, and Haber, 1970), that an image, if well formed, should persist regardless of its bizarreness. Thus it is possible that the subjects' bizarre images for some reason may have been less complete, less detailed, or less 'vivid.' For all groups, the 30-sec maximum for image construction was usually ample; nevertheless, the subjects presented with the bizarre triplets may have adopted a less stringent criterion for vividness or completeness.

A second possibility for the superiority of plausible imagery is that a common triplet may easily suggest several images to a subject, whereas a bizarre triplet suggests only one. A greater number of images leading to a given response could increase the likelihood of that response at the time of recall.

Figural unity, which Atwood (1969) has shown to be important for remembering an image, probably did not play a role in the present experiment. The subjects in both imagery conditions were instructed to create an image that linked the two nouns in some manner specified by the verb, so that there is no reason to believe that the parts of a bizarre image were less spatially interdependent than those of a common image.

One final point concerns what is meant by 'common' as opposed to 'bizarre' images as mnemonic aids. One might object that there is nothing particularly common about ELEPHANT TRAMPLE GARDEN or QUEEN PLAY HARE (two of the common triplets used); that is, that the triplets do not truly span a continuum from common to bizarre. In fact, 'commonness' is probably an inappropriate term when used in the context of mnemonic imagery. A truly common image in a practical mnemonic system would be an extremely rare event, since the essence of such a system is the pairing of the object to be remembered with one of an arbitrary list of items, a pairing chosen simply because the items are easy to remember in a particular order. It therefore may be more reasonable to refer to a continuum running from 'plausible' to 'bizarre.'

In conclusion, and aside from the particular reasons for the superiority of common over bizarre imagery, it seems fair to say that any practical advice to a person for using imagery as a mnemonic system should, while stressing concreteness and figural unity, direct him to imagine the most plausible connection between the retrieval cue and the object to be remembered, and not to waste effort constructing an especially bizarre, implausible, or unusual image.

Notes

This research was performed under Contract N00014-67-A-0163-0001 between the Johns Hopkins University and the Psychological Psychology Branch, Office of Naval Research. Dr. Collyer's address is now: Technology Incorporated, Life Sciences Division, 8531 North New Brauntfels Avenue, San Antonio, Texas 78217. Mr. Jonides's address is now: Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104. Received for publication April 23, 1971.

1. Some examples of nouns rejected on this basis would be those in brackets: HORSE PULL [CARRIAGE], SHOTGUN BLAST [THIEF], OF TABLESPOON MEASURE [SUGAR].

2. Response nouns rejected on this basis might have been (in brackets) HORSE PULL [STRAWBERRY], SHOTGUN BLAST [FORK], OF TABLESPOON MEASURE [HEADLIGHT].

References

- Atwood, C. E. 1969. Experimental studies of mnemonic visualization. Unpublished doctoral dissertation, University of Oregon.
- Briggs, G. C., Hawkins, S., and Crovitz, H. F. 1970. Bizarre images in artificial memory. *Psychonomic Science* 19:353-354.
- Bugelski, B. R. 1968. Images as mediators in one-trial paired associates learning: II. Self-timing in successive lists. *Journal of Experimental Psychology* 77:328-334.
- Bugelski, B. R. 1970. Words and things and images. Presidential address, Eastern Psychological Association, Atlantic City, April.
- Bugelski, B. R., Kidd, E., and Segmen, J. 1968. Image as a mediator in one-trial paired-associate learning. *Journal of Experimental Psychology* 76:69-73.
- Kirk, R. E. 1968. *Experimental design: Procedures for the behavioral sciences*. Belmont, Calif.: Wadsworth Publishing.
- Miller, G. A., Galanter, E., and Pribram, K. H. 1960. *Plans and the structure of behavior*. New York: Holt.
- Paivio, A. 1969. Mental imagery in associative learning and memory. *Psychological Review* 76:241-263.
- Paivio, A., Yuille, J. C., and Madigan, S. A. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology* 76(1, part 2).
- Standing, L., Conezio, J., and Haber, R. N. 1970. Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science* 19:73-74.
- Wallace, W. H., Turner, S. H., and Perkins, C. C. 1957. Preliminary studies of human information storage. University of Pennsylvania Institute of Cooperative Research, Signal Corps project 1326.
- Wood, C. 1967. Mnemonic systems in recall. *Journal of Educational Psychology* 58(6, part 2).
- Yates, F. A. 1966. *The art of memory*. London: Routledge & Kegan Paul.