

A reassessment of typicality effects in free recall

PAUL WHITNEY, THOMAS G. COCKLIN, JAMES F. JUOLA, and GEORGE KELLAS
University of Kansas, Lawrence, Kansas

In a recent study, Greenberg and Bjorklund (1978) reported data that they interpreted as supporting a differential-category-encoding account of typicality effects. In the present article, we point out several studies that contradict the conclusions drawn by Greenberg and Bjorklund. In addition, we present a replication of their study, but with a greater manipulation of typicality. The data do not support the differential-encoding hypothesis. It was concluded that the available evidence supports a feature-overlap account of typicality effects.

A particularly influential approach to the study of semantic memory has been taken by Rosch (1975a). She proposed that natural categories have an internal structure in which membership is based on degree of similarity to the prototype, or central tendency, of the category. In accord with this view, Rosch (1975a) found that priming with the superordinate category name facilitated category judgments for physically identical, typical exemplars but inhibited judgments for atypical exemplars. According to Rosch and Mervis (1975), subjects process typical and atypical exemplars differently in a categorization task because typical exemplars have more features in common with the category prototype.

In contrast, Loftus (1975) proposed that the typical and atypical items are categorized differently because the atypical items are often salient members of other categories. In her reply to Loftus, Rosch (1975b) noted that when atypical exemplars were actually replaced by members of other categories, the pattern of results did not match the pattern that resulted when atypical exemplars were employed. In addition, using a release-from-proactive-interference (PI) task, Keller and Kellas (1978) found similar rates of PI buildup for typical and atypical conditions. This result indicates that atypical exemplars were not differentially encoded with regard to category membership.

In opposition to Rosch's theory, Greenberg and Bjorklund (1981) obtained evidence that they interpreted as strongly supporting Loftus's (1975) dif-

ferential-category-encoding account of typicality. They presented subjects with either typical or atypical category exemplars to be memorized in a three-trial, free-recall task. Half the subjects received the category name with each exemplar, and half received only the exemplars. Greenberg and Bjorklund reasoned that if the differential-encoding hypothesis is correct, recall and clustering scores should be higher for typical than for atypical exemplars when the category names are not provided. When category names are given at input, the atypical exemplars should be uniformly encoded, which should yield recall and clustering scores as high as those for typical exemplars. The results were in agreement with these predictions.

Although Greenberg and Bjorklund's (1981) data seem to favor the differential-encoding hypothesis, there are problems with the interpretation of their data. Greenberg and Bjorklund attempted to vary typicality and to control for production frequency. Since typicality and production frequency are highly correlated means of assessing the internal structure of semantic categories (Mervis, Rosch, & Catlin, 1976), it is difficult to manipulate typicality only. Because of this, Greenberg and Bjorklund could make only a small manipulation of typicality: The mean ratings of the typical and atypical exemplars were 1.74 and 3.38, respectively (on Rosch's 7-point scale, with 7 being highly atypical). Some examples of their "atypical" exemplars include *lamp* and *bench* as furniture and *vest* and *scarf* as clothing. It seems unlikely that subjects would assign such exemplars to alternative categories.

Given the problems with the stimuli used by Greenberg and Bjorklund (1981), we replicated their procedure but used exemplars more clearly atypical of their respective categories. It was predicted that under these conditions we would find recall and clustering differences between typical and atypical exemplars even if category names were provided at input.

This research was partially supported by U.S. Public Health Service Grant HD 00870 from the National Institute of Child Health and Human Development to the Kansas Center for Mental Retardation and Human Development. The authors wish to thank Mary McQuaid and Susan Kemper for their valuable service and timely suggestions on an earlier version of this paper. Address reprint requests to Paul Whitney, Department of Psychology, 426 Fraser Hall, University of Kansas, Lawrence, Kansas 66045.

METHOD

Subjects

The subjects were 32 native English-speaking students enrolled in an introductory psychology class at the University of Kansas. All received partial course credit for their participation.

Materials

The stimuli were selected from the Uyeda and Mandler (1980) typicality norms. A total of 36 typical and 36 atypical items were sampled, 6 from each of the categories used by Greenberg and Bjorklund (1981): furniture, clothing, sports, vehicles, weapons, and birds. The typical and atypical exemplars were balanced for Thorndike-Lorge (1944) word frequencies. The mean typicality ratings of the typical and atypical exemplars were 1.83 and 4.29, respectively. Each target stimulus was typed on a 3 × 5 in. card. For the category-at-input conditions, the category name was also typed on the card above the exemplar and enclosed in parentheses. Thus, four sets of stimuli were prepared: typical exemplars only, atypical exemplars only, typical exemplars with category names, and atypical exemplars with category names.

Design and Procedure

The design and procedure followed those of Greenberg and Bjorklund's (1981) study. There were two between-subjects factors, input condition (category names, no category names) and typicality (typical, atypical). The single within-subjects factor was trials.

All subjects were tested individually in a single, 20-min session. Each subject was assigned to one of the four stimulus groups. The subjects in the category-at-input condition were told to look at the word in parentheses because it would help them to remember the target word. All subjects were told to try to recall as many of the target words as possible. The cards were presented in random order at a rate of 4 sec per card. After the last item, the subjects performed a 30-sec interpolated counting task that was followed by a 3-min free-recall period. During the recall period, the stimulus set was reshuffled into a new random order. The second and third trials for each subject used the same procedure and stimulus set.

RESULTS AND DISCUSSION

The mean recall and clustering scores on each trial are shown in Table 1. The clustering measure used was the ratio of repetitions (Bousfield, 1953). The alpha error level for all tests was set equal to .05.

An analysis of variance on the recall scores revealed a significant increase in recall over trials [$F(2,56) =$

249.8, $MSe = 4.8$], which was qualified by the typicality × trials interaction [$F(2,56) = 7.6$, $MSe = 4.8$]. Tukey B tests of this interaction showed that typical exemplars were recalled significantly better than were atypical exemplars on the first two trials but not on the third. It should be noted that several subjects in each group had perfect recall scores on the third trial, which contributed to an overall ceiling effect. Most notably, the interaction of typicality and input condition obtained by Greenberg and Bjorklund (1981) did not even approach significance in the present study ($F < 1$). The differences in recall between typical and atypical exemplars on the first two trials were statistically equivalent across both input conditions. This is contrary to the differential-category-encoding hypothesis, which maintains that recall differences between typical and atypical exemplars should be eliminated when category names are provided at input. Our results are consistent with the feature-overlap hypothesis (Rosch & Mervis, 1975), which predicts that typicality effects should be consistent across both input conditions because recall will be based on the amount of feature overlap among exemplars. Thus, atypical words are recalled less well, not because they have ambiguous relations with the categorical structures of the list, but because they are not as closely integrated with the central tendency of the category. Atypical exemplars, then, are less well recalled whether category labels are given or not, due to their relative unavailability during a search of the category.

The analysis of the clustering scores is equally damaging to the differential-encoding hypothesis. This analysis revealed greater clustering for typical than for atypical exemplars [$F(1,28) = 10.9$, $MSe = 5.7$] and an increase in clustering over trials [$F(2,56) = 170.9$, $MSe = .05$]. None of the interactions reached significance. Again, the advantage of typical over atypical exemplars was statistically equivalent over input conditions. This contrasts sharply with Greenberg and Bjorklund's (1981) results. It is possible that in their study the small typicality effects

Table 1
Mean Recall and Clustering by Input Condition, Typicality, and Trials

| | Trial | | | | | | Overall Mean | |
|------------------|-------|-----|-------|-----|-------|-----|--------------|-----|
| | 1 | | 2 | | 3 | | Mean | C |
| | Mean | C | Mean | C | Mean | C | | |
| No Category Name | | | | | | | | |
| Typical | 22.75 | .43 | 28.62 | .63 | 31.13 | .69 | 27.50 | .58 |
| Atypical | 17.00 | .23 | 26.62 | .41 | 31.50 | .53 | 25.04 | .39 |
| Category Name | | | | | | | | |
| Typical | 22.00 | .48 | 30.12 | .68 | 33.25 | .77 | 28.46 | .64 |
| Atypical | 18.00 | .31 | 27.37 | .53 | 31.62 | .69 | 25.66 | .51 |

Note—C = clustering score.

they obtained when no category names were provided were masked by the overall improvement in recall and clustering when category names were provided.

A possible alternative explanation of the present results is that the typical exemplars were better remembered than were atypical exemplars in the category-at-input condition because the typical exemplars were higher than the atypical exemplars in production frequency (PF). If so, PF should be an important determinant of recall when the effects of typicality are controlled. To examine this possibility, the typical exemplars were divided into high and low PF groups based on the Battig and Montague (1969) norms. The PF of the typical exemplars ranged from 440 to 25, with a median of 234.5. Exemplars were designated as high or low in PF by whether they were above or below the median PF. This same procedure was followed in dividing the atypical exemplars into two levels of PF. The atypical exemplars ranged from 420 to 12, with a median of 36. A PF x trials analysis of variance was performed on the recall scores of the subjects who received category names and typical exemplars. The only significant effect in this analysis was that of trials [$F(2,14) = 23.8$, $MSe = 5.2$]. The PF x trials analysis of recall scores of subjects who received category names and atypical exemplars revealed the same results. Again, the only significant effect was trials [$F(2,14) = 66.8$, $MSe = 3.26$]. Thus, with typicality controlled (i.e., all typical or atypical exemplars), PF did not affect recall even though PF varied widely within each stimulus group.

In summary, even when category names are provided, typical exemplars are better recalled and more highly clustered than are atypical exemplars. These results provide converging evidence for several lines of previous research that have indicated that differ-

ential category encoding cannot serve as a complete explanation for typicality effects.

REFERENCES

- BATTIG, W., & MONTAGUE, W. Category norms for verbal items in 56 categories. *Journal of Experimental Psychology Monograph*, 1969, **80**(3, Pt.2).
- BOUSFIELD, W. A. The occurrence of clustering in recall of randomly arranged associates. *Journal of General Psychology*, 1953, **49**, 229-240.
- GREENBERG, M. S., & BJORKLUND, D. F. Category typicality in free recall: Effects of feature overlap or differential category encoding? *Journal of Experimental Psychology: Human Learning and Memory*, 1981, **7**, 145-147.
- KELLER, D., & KELLAS, G. Typicality as a dimension of encoding. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, **4**, 78-85.
- LOFTUS, E. Spreading activation within semantic categories: Comments on Rosch's "Cognitive representations of semantic categories." *Journal of Experimental Psychology: General*, 1975, **104**, 234-240.
- MERVIS, C., ROSCH, E., & CATLIN, J. Relationships among goodness of example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, 1976, **7**, 283-284.
- ROSCH, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 1975, **104**, 192-233. (a)
- ROSCH, E. Reply to Loftus. *Journal of Experimental Psychology: General*, 1975, **104**, 241-243. (b)
- ROSCH, E., & MERVIS, C. B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 1975, **7**, 573-605.
- THORNDIKE, E., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teacher's College Press, Columbia University, 1944.
- UYEDA, K. M., & MANDLER, G. Prototypicality norms for 28 semantic categories. *Behavior Research Methods & Instrumentation*, 1980, **12**, 587-595.

(Manuscript received for publication July 1, 1983.)