Literal versus nonliteral reminders for proverbs

JON G. TEMPLE and RICHARD P. HONECK
University of Cincinnati, Cincinnati, Ohio

Previous research has indicated a relationship between the degree of figurative understanding and the effectiveness of different kinds of recall prompts for proverbs. In this study, subjects wrote interpretations of unfamiliar proverbs. On a subsequent reminding task, the subjects were presented several different types of prompts for recall of the proverbs; the prompts exhibited different kinds of similarity relationships with the proverbs. Although it was expected that higher levels of understanding, as indexed by subjects' interpretations, would make the more abstract prompts more effective, this was not found. It was suggested that subjects who did not have a good understanding of the proverbs were able to use the abstract prompts to recall the proverbs via a compensatory recall strategy.

The most important determinant of reminding is often thought to be the similarity between a recall cue and the thing learned. However, recent research suggests that there are different kinds of similarity and that they may have differential effects on reminding. For example, research by Gick and Holyoak (1983) and Novick (1988) has shown that people have difficulty recognizing and retrieving analogically similar items in memory. Also, Gentner (1983; Genter & Landers, 1985) have concluded that analogical reminding is better determined by surface or literal similarity than by more deeply conceptual analogical similarity.

Other research has suggested that not all abstract or nonliteral prompts are equally useful in reminding. Dorfmueller and Honeck (1980) examined linguistic families consisting of a proverb, an interpretation of the proverb, and two sentence-length instances that illustrated the figurative meaning of the proverb. On a recall task, they found that performance was best when the interpretation was used as a recall prompt for other family members; other family members were poorer prompts. Dorfmueller and Honeck suggested that recall was facilitated by the formation of an abstract, nonlinguistic, nonimagistic "conceptual base" or "miniature theory" for family membership. The interpretations, being expressed in abstract, general terms, more nearly reflected this conceptual base than did other family members and consequently served as a superior recall prompt.

In a study that was similar to the present one, Schumacher (1987) solicited recall of proverbs using both surface and nonliteral cues. Half of the subjects rated 50 proverbs on cleverness ("superficial" encoding), and the other half wrote interpretations of the proverbs ("deeper" encoding). Ten pairs of these proverbs were analogically related (i.e., their figurative meanings were similar), another 10 pairs were only superficially related (i.e., they had content word

overlap), and the remaining 10 proverbs were unrelated fillers. On a subsequent reminding task, surface-similar prompts (i.e., proverbs that shared content words, but not the figurative meanings, of the original proverbs) proved to be more effective for retrieval of the proverbs than analogically similar prompts (i.e., proverbs that had the same figurative meanings as the original proverbs). However, proverbs that were prompted by analogically similar cues were recalled significantly more often if deeper encoding had occurred than if they were superficially encoded. Moreover, type of encoding did not influence recall of the proverbs that were prompted by surface-similar cues.

Because of limitations in Schumacher's (1987) methodology, several important questions about the determinants of reminding were not adequately addressed. For example, it is not clear what precisely was being recalled, because only one keyword, not an entire proverb, was sufficient to meet the recall criterion. In addition, both surface- and analogically related pairs of proverbs were presented during acquisition. Surface-similar proverb pairs were not prompted by analogically similar recall prompts, or vice versa. It is possible that both the surface and analogical connections were previously established on the basis of the proverb pair and that the prompt simply triggered recognition of this preexisting relationship. As a consequence, the paradigm might be better described as involving recognition than recall, leaving the effects of encoding on cued recall an open question. Furthermore, Schumacher did not look at the degree of interpretation goodness. There is reason to believe from other studies (Honeck & Firment, 1989; Honeck, Riechmann, & Hoffman, 1975) that degree of interpretation goodness, as an index of a subject's understanding, may be used to predict retrieval of analogically similar items.

The present study examined the relationship between the degree of figurative understanding and the effectiveness of different kinds of prompts for proverb recall. We used several prompts that varied in semantic distance from the proverb in a two-phase experiment. During the learning phase, one group received proverbs along with in-

---

stances of the proverbs and a second group received proverbs without instances. This context manipulation was designed to promote better figurative understanding by the context group. All subjects wrote interpretations of the proverbs. Proverb recall was cued with prompts that exhibited different kinds of similarity relationships with the proverbs. We expected that higher levels of understanding, as indexed by subjects' interpretations, would make the more abstract prompts more effective.

METHOD

Subjects
Forty-eight subjects from the University of Cincinnati participated.

Materials
Twelve sets of stimuli were constructed. Each stimulus set consisted of a proverb (the *old* proverb), its interpretation, a subject noun and a theme word derived from the old proverb, a *new* proverb similar in meaning to the old proverb, a sentence-length instance of the figurative meaning of the old proverb (the *new* instance), a longer instance of the figurative meaning of the old proverb (the *old* instance), and finally, a *mere-appearance* proverb that incorporated the same subject noun as the old proverb but that had a different figurative meaning. For example, one old proverb was *"Bees have honey in their mouths and stingers in their tails."* For this old proverb, the rest of the stimulus set was as follows: interpretation (*"Some things that look good on the surface may turn out to be totally bad"*), subject noun (*"bees"*), theme word (*"deceptiveness"*), new (semantically related) proverb (*"The friendliest cats can have the sharpest claws"*), new instance (*"The loan shark was more than happy to give the indebted gambler $5,000 at 25% interest"*), old instance (*"The psychopath had a taste for excellent champagne, fine caviar, and unusual gourmet foods, and he liked to share them with beautiful women. He drove through the city and lured women into his fancy car. Then he took them home, fed them one of his favorite meals, and killed them"*), and mere-appearance (surface-related) proverb (*"In the hive, the queen bee reigns supreme"*).

The old proverbs were the to-be-recalled items. The old instance served as context for the context group. The other materials were all used as recall prompts, with the subject nouns and the mere-appearance proverbs serving as surface-level prompts and the new proverbs, new instances, interpretations, and theme words serving as more abstract reminders of the old proverbs. In addition, the interpretations were used as a referent for judging the goodness of subjects' interpretations.

The proverbs and instances were unfamiliar to most subjects and required no special knowledge or vocabulary to be understood. An independent group of 31 subjects rated the old proverbs for familiarity. On a 7-point scale on which 7 = very familiar, the mean rating was 2.57. With the exception of the surface-level prompts, no overlap of content words existed between any of the materials, and semantic similarity between stimulus sets was kept to a minimum.

An independent group of 25 subjects rated the semantic similarity between the new proverbs and the old proverbs. On a 9-point scale on which 9 = very similar, the mean rating was 6.50. An independent group of 22 subjects rated the semantic similarity between the remaining abstract materials and the old proverbs. On a 5-point scale on which 5 = very similar, the new instances, the old instances, the interpretation, and the theme word had mean ratings of 3.87, 3.98, 4.30, and 3.66, respectively.

Design
A mixed design was used, with one between-subject factor (context bias: context or no context) and one within-subject factor (type of recall prompt: subject noun, theme word, new (semantically related) proverb, mere-appearance proverb, new instance, or interpretation). The subjects in the context group received the old proverb along with an old instance, while the subjects in the no-context group were presented each proverb by itself.

Procedure
The subjects were tested in groups of 6-12 in a quiet room.

Acquisition phase. The subjects were given a 13-page booklet. The first page instructed subjects to write an interpretation for each of the 12 old proverbs (the term "proverb" was used) on the following pages, 1 proverb per page. In addition, the context group received an instance presented above each proverb and had modified instructions that specified the general relationship between the instance and the proverb. Written examples of both a proverb and its interpretation were provided for each group. The context group also received an example of an instance. After the subjects read the instruction page, the experimenter orally delivered the instructions and then answered any questions that arose.

The subjects were permitted 2 min to read the materials and to write an interpretation for each proverb. The subjects were not permitted to go back to previous pages. After 24 min, the booklets were collected and the recall phase began.

Recall phase. The subjects were given a second 13-page booklet. The first page provided examples of the six recall prompt types, with instructions to think about the relationship between each of the prompt types and a corresponding old proverb. The subjects were also informed that they would be given one of these prompt types on each of the remaining pages of the booklet and that they were to use these prompts to recall verbatim the proverbs for which they had written interpretations. The experimenter answered any questions that arose and then orally delivered the instructions. Recall was self-paced.

An experimental session was completed within 50 min.

RESULTS

Acquisition Phase
The subjects' interpretations were rated on a 9-point scale of goodness, with 9 = very good. Interrater reliability (r_{xx}) of the experimenter and an associate familiar with the materials ($n = 48$ per sample) was .77, .81, .89, and .96 ($M = .86$) for the "tree," "tie," "lion," and "fish" sets of interpretations, respectively. The ratings of the interpretations by subjects in the context group ($M = 4.68, SD = .959$) and by subjects in the no-context group ($M = 4.24, SD = 1.15$) did not differ significantly [$t(46) = 1.77, p > .05$, two-tailed].

Recall Phase
Two measures of recall performance were taken—the percentage of content words recalled in any order and the percentage of content words recalled in correct order. On a prompt-by-prompt basis, the number of content words recalled in correct order was 90% of the number of content words recalled in any order. Since verbatim recall was the focus of this study, only those analyses performed with the correct order measure will be reported.

The mean percentage of content words recalled was subject noun ($M = 71\%, SD = .20$), mere-appearance proverb ($M = 51\%, SD = .33$), new proverb ($M = 46\%, SD = .34$), interpretation ($M = 45\%, SD = .31$), new instance ($M = 39\%, SD = .27$), and theme word ($M = 26\%, SD = .25$) (see Figure 1).

A one-between (context bias: context or no context), one-within (prompt type) analysis of variance was performed on the arcsine transformed percentage recall measure. There was no significant effect of context bias [$F(1,46) < 1, MS_e = .431, p > .05$] or of the interaction of context bias and prompt type [$F(5,230) < 1, MS_e = .301, p >$
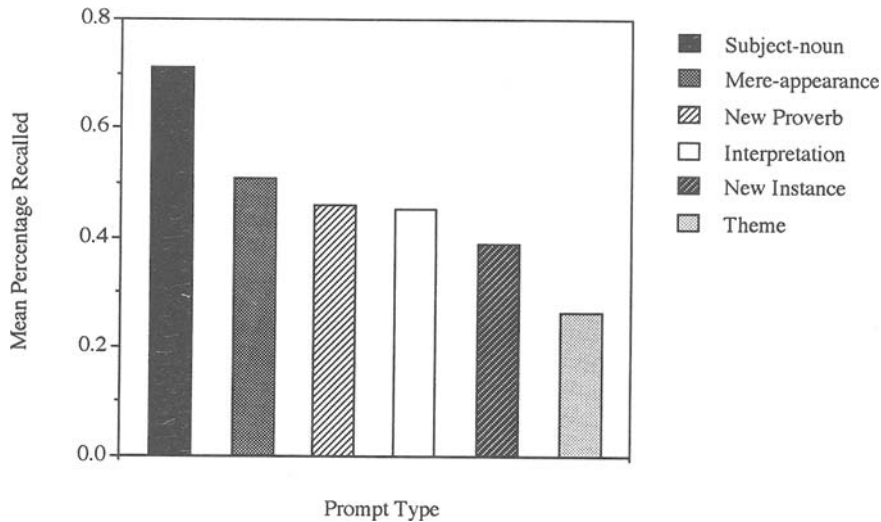
Figure 1. Mean percentage of content words recalled as a function of prompt type.

.05]. However, prompt type was significant [$F(5,230) = 12.141, MS_e = 7.599, p < .0001$]. A Newman-Keuls test indicated ($\alpha = .05$) that the subject noun served as the most effective recall prompt and that the theme word served as the least effective recall prompt. The mere-appearance proverb, new instance, new proverb, and interpretation prompts were equally effective.

Relationship Between Interpretations and Prompt Effectiveness

A major purpose of the present study was to examine the relationship between prompt effectiveness and interpretation goodness. The percentage of content words recalled was correlated with interpretation ratings for each subject per old proverb. The correlations were calculated for each prompt type. None of the overall Pearson correlations were significant.

DISCUSSION

It was found that the subject noun, a surface-level prompt, was more effective than the abstract prompts. However, not all surface-level prompts were alike. The subject-noun prompt was more effective alone than when embedded in a short context, the mere-appearance proverb. The abstract prompts—the new proverb, interpretation, and new instance—were equally effective. Nevertheless, the subjects were able to use the theme-word prompt, the most abstract and least effective prompt, to recall 26% of the words in correct order, a surprisingly high level of recall considering the degree of abstraction. The absolute level of recall for the abstract prompts was higher than was found in similar studies. For example, Gentner and Landers (1985) reported about 20% recall, and Rattermann and Gentner (1987) reported less than 10% recall when subjects used a nonliteral prompt.

Although other research suggested that there might be a relationship between interpretation goodness and degree of recall (Gick & Holyoak, 1983; Honeck et al., 1975; Schumacher, 1987), this was not found in the present study. Most of the correlations were nonsignificant, or typically small and negative when significant. This was surprising, partic-

ularly because the subjects were clearly able to use two of the abstract prompts to recall at about a 46% rate. Nevertheless, there were important methodological differences between the present study and earlier studies that may shed light on this discrepancy. For example, Schumacher (1987) failed to investigate the effects of recall independently of recognition, thereby permitting his subjects to concentrate more on the nonliteral meaning. Verbal ability may also have played a role—some subjects may have understood the proverb but failed to put that understanding into a good explicit form, thereby curtailing the range of interpretation.

Another possible reason why the interpretation/recall correlation was nonsignificant is that those subjects who did not write good interpretations, and presumably did not have a good understanding of the acquisition proverbs, were able to use the abstract prompts to recall the acquisition proverbs via a compensatory recall strategy. Because only 12 concrete proverbs had to be kept in memory and the possible number of interpretations of any single proverb was also small, subjects could mentally scan through their list of acquisition proverbs and interpret each in terms of the nonliteral prompt. When a nonliteral match was found, subjects could then eliminate a possibility from their mental list when encountering the next nonliteral prompt. Essentially, the subjects *had* to find a nonliteral connection between a proverb and a prompt in order to use the prompt—if the relationship was not acquired during learning, then it had to be constructed during recall.

We suggest that future research should therefore be careful to address the possibility of such a strategy being used. The effect of this strategy could be attenuated by increasing the number of to-be-recalled items, by setting a time limit on recall, and by prohibiting subjects from skipping around in the test booklet.

REFERENCES

DORFMUELLER, M. A., & HONECK, R. P. (1980). Centrality and generativity within a linguistic family: Toward a conceptual base theory of groups. *Psychological Record*, **30**, 95-109.

GENTNER, D. (1983). Structure-mapping: A theoretical framework. *Cognitive Science*, **7**, 155-170.

GENTNER, D., & LANDERS, R. (1985). Analogical reminding: A good match is hard to find. *Proceedings of the International Conference on Systems, Man, and Cybernetics*. Tucson, AZ.

GICK, M. L., & HOLYOAK, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, **15**, 1-38.

HONECK, R. P., & FIRMENT, M. (1989). Accessing abstract categories. *Bulletin of the Psychonomic Society*, **27**, 206-208.

HONECK, R. P., RIECHMANN, P., & HOFFMAN, R. R. (1975). Semantic memory for metaphor: The conceptual base hypothesis. *Memory & Cognition, 3*, 409-415.

NOVICK, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14*, 510-520.

RATTERMANN, M. J., & GENTNER, D. (1987). Analogy and similarity: Determinants of accessibility and inferential soundness. *Proceedings of the Ninth Annual Meeting of the Cognitive Science Society* (pp. 23-34), Seattle.

SCHUMACHER, R. M. (1987). *Similarity-based remindings: Effects of distance and encoding on retrieval.* Unpublished master's thesis, University of Illinois, Urbana-Champaign, IL.