

# On memory for metaphor

MARC MARSCHARK and R. REED HUNT

*University of North Carolina at Greensboro, Greensboro, North Carolina*

Three experiments investigated how metaphors are represented in memory and the effects on memory of variables known to affect metaphor comprehension. Ten theoretically relevant dimensions were examined. In Experiments 1 and 2, free recall across a variety of orienting tasks was consistently and positively predicted by the rated imageability of metaphoric topics (sentence subjects) and the rated number of interpretations for each metaphor. The number of interpretations effect was reversed in the cued recall task of Experiment 3, as metaphors with fewer interpretations were remembered better. These and other results supported previous suggestions that memory for metaphor involves wholistic representations akin to metaphoric grounds. They also revealed consistent differences in the roles of several variables in comprehension and memory for metaphor.

Metaphor is a challenge to psychologists. As one of the more enigmatic problems in language research, it raises two fundamental questions: How are these literally false assertions understood as acceptable? And how are they mentally represented? Although some progress has been made in understanding how metaphors are comprehended (e.g., Gildea & Glucksberg, 1983; Glucksberg, Gildea, & Bookin, 1982; Tourangeau & Sternberg, 1981; see also collections in Honeck & Hoffman, 1980; Ortony, 1979), the study of metaphor remains largely outside the mainstream of cognitive investigation. This isolation is likely due to the fact that theories of metaphor comprehension have not overlapped conceptually with more general theories of language (H. Clark & E. Clark, 1977; Paivio & Begg, 1981). Many of the variables deemed important by metaphor researchers seem specific to figurative language, and few attempts have been made to integrate findings from metaphor studies into other areas of cognitive psychology. The present experiments, therefore, were designed to relate metaphor research to more traditional cognitive investigation while expanding current knowledge about both figurative language and memory in general.

Most previous studies of metaphor have examined figurative comprehension or judgments of other metaphoric attributes. Several variables have been found to affect ease of interpretation and metaphor "goodness," but possible relationships among these variables have not been explored beyond those predicted for specific pairs of vari-

ables by particular theories (see Marschark, Katz, & Paivio, 1983, for a review). Recently, however, Marschark et al. (1983; see also, Katz, Paivio, & Marschark, 1985) conducted a normative study of 10 attribute dimensions of current theoretical interest in metaphor research. In two experiments, subjects rated a pool of novel metaphors on 10 dimensions: felt familiarity, comprehensibility, ease of interpretation, estimated number of alternative interpretations, metaphor goodness, degree of metaphoricity, sentence imageability, imageability of topic (subject), imageability of vehicle (predicate), and the semantic relatedness of topic and vehicle. Marschark et al. (1983) found mixed support for several current theoretical positions concerning metaphor comprehension, each having strong and weak areas of applicability. Perhaps of more general interest, both the Marschark et al. (1983) study with invented metaphors and the Katz et al. (1985) study with selected poetic metaphors revealed all 10 dimensions to be highly interrelated. This finding suggests caution in generalizing from earlier studies that have involved metaphors selected on the basis of only one or two dimensions and indicates the need for more controlled investigation of metaphors in comprehension as well as other tasks.

Given the abundance of recent metaphor research, surprisingly few studies have examined memory for metaphor. Among these few, however, there is general agreement that memory for metaphor is based on some wholistic, conceptual format transcending simple verbal and imaginal associations between topic and vehicle nouns (cf. Malgady & Johnson, 1976, 1980; Tourangeau & Sternberg, 1981). Verbrugge and McCarrell (1977), for example, suggested that metaphoric grounds (i.e., similarities between topic and vehicle) serve as the functional units of memory for metaphor. Their subjects listened to a series of metaphors and then were tested for recall using the topics, vehicles, or grounds as prompts. Verbrugge and McCarrell found that all three cue types were effective in facilitating recall, leading them to conclude that

Portions of this research were supported by Grant 1-RO1-NS20064-01 to Marc Marschark from the National Institute of Neurological and Communicative Disorders and Stroke and a grant from the Research Council of the University of North Carolina at Greensboro. The authors thank Michael Penland for his assistance in Experiments 2 and 3, Kim Young (Dunnabart) and Monica Wallace for conducting Experiment 1, and Ian Begg for his many helpful comments on an earlier version of this paper.

Correspondence should be sent to Marc Marschark, Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC 27412.

the memorial representation of a metaphor involves an abstract resemblance between the topic and vehicle domains rather than "the recognition of preexisting attributes associated with topics, the priming or weighting of salient attributes during acquisition, or the transfer of salient attributes associated with vehicles" (Verbrugge & McCarrell, 1977, p. 525).

Honeck (1973) also argued for semantic representation in memory for figurative language, and he contrasted it with a representation based on the grammatical structure of the input sentence. He had college students learn a series of proverbs, each accompanied by either its repetition, a grammatical transformation, a paraphrase, or a control sentence. The students, given the subject nouns as cues, then attempted to recall the proverbs. Honeck found that, with both high- and low-imagery proverbs, recall was significantly higher if the items had been learned in the context of their paraphrases than if they had been learned in the context of any of the other types of sentence. The failure to obtain an interaction between input condition and proverb imageability led him to suggest that imagery could not be primarily responsible for memory for figurative language. Honeck suggested instead that some abstract form of storage was more likely.

Reichmann and Coste (1980) made a similar argument in their study of imagery and proverbs. They cited evidence showing images of a proverb's literal reading to interfere with comprehension of its intended (figurative) meaning. Empirically, Reichmann and Coste failed to obtain reliable correlations between proverb imagery and degree of metaphoricality (cf. Marschark et al., 1983; Katz et al., 1985) or imagery and various recall measures, leading them to conclude that imagery is only tangentially related to metaphor comprehension "or, at best, a process essential only at a shallow level of processing" (Reichmann & Coste, 1980, p. 199).

With regard to both Reichmann and Coste's (1980) and Honeck's (1973) results, we note that no one as yet has compared memory for proverbs and memory for metaphors. Although most researchers of proverb comprehension and memory cautiously generalize their results to metaphors, there are some variables on which coincidence is less than likely. Imagery is one example. A primary attribute of proverbs is that their intended meanings only incidentally relate to their literally true surface readings. The latter usually are but one instantiation of the proverbial point (which must be general by definition). Therefore, generating an image, for example, of the act of sewing will no more facilitate understanding of *A stitch in time saves nine*, than an image of equine molars will for *Don't look a gift horse in the mouth*. Metaphors, however, are quite different. The value of literally false statements such as *Ants are bulldozers* or *Islands are corks* may lie precisely in perceptual or imagined similarities that allow and sometimes require more general conclusions (Ortony, 1975). Honeck's (1973) and Reichmann and Coste's (1980) failures to obtain statistical interactions between various instructions and rated imagery, in any case, should

not have been interpreted as indicating imagery to be epiphenomenal in figurative language, because main effects of imageability were obtained in both studies.

## EXPERIMENT 1

Given how little is known about variables that might influence memory for metaphors, previous studies involving manipulation of only one or two variables are difficult to interpret. Consequently, in the first experiment we empirically examined the effects on memory for metaphor of variables previously identified as important for metaphor comprehension or for memory for literal language. The assessment of 10 different normative dimensions allowed for greater experimental control than previously has been possible.

Ten groups of subjects rated the same set of metaphors, each on a different one of the 10 dimensions examined by Marschark et al. (1983) and Katz et al. (1985). Relationships between the normative ratings and subsequent incidental free recall of the metaphors then were examined. The goal was to determine which, if any, of the variables would predict recall of the metaphors.

### Method

**Stimuli, design, and procedure.** The stimuli were 98 metaphors used by Marschark et al. (1983, Study II). All were in the form "(An) A is (a) B" or "As are Bs." Table 1 lists the mean ratings and standard deviations of the stimuli on each of 10 normative dimensions. Except for number of interpretations, all are based on 7-point scales from "low" to "high." The stimuli were printed on four pages of a test booklet, and each booklet was randomly combined with one of 10 different cover pages containing instructions and several examples. There were thus 10 different orienting conditions, each requiring subjects to rate the 98 metaphors on one of the 10 dimensions listed in Table 1. The instructions were identical to those used by Marschark et al. (1983) and Katz et al. (1985). Appendix A provides complete instructions for the number of interpretations condition and excerpts from the instructions for the other conditions. Complete sets of the stimuli, norms, and instructions are available from the authors.

Each subject read a set of instructions and was given 20 min to do the self-paced rating tests. Subjects were told not to look back over the booklets if they finished before the allotted time expired.

Table 1  
Mean Normative Ratings and Recall of  
98 Metaphors Used in Experiment 1

Normative Dimension	Mean Rating	Standard Deviation of Ratings	% Recall
1. Number of Interpretations (INT)	1.81	0.26	10
2. Ease of Interpretation (EOI)	5.08	1.28	9
3. Degree of Metaphoricity (DOM)	3.41	0.79	8
4. Metaphor Goodness (GOOD)	4.30	1.04	8
5. Imagery (I)	4.56	1.25	8
6. Imagery of Subject (IS)	4.66	1.24	9
7. Imagery of Predicate (IP)	4.47	1.03	9
8. Semantic Relatedness (SEM)	3.94	1.28	12
9. Felt Familiarity (FAM)	3.94	1.20	9
10. Comprehensibility (COMP)	4.97	1.04	9

Note—All means based on 7-point scales except for number of interpretations.

Following the rating task, test booklets were collected, and each subject was given a blank sheet of paper. The experimenter then orally gave free-recall instructions. Subjects were asked to use original wording if possible but otherwise to write any ideas or parts of items they could not remember fully. In the first four testing sessions, subjects were given as much time as they needed to complete the recall task and were queried several times as to their progress. All subjects finished within 10 min. In subsequent sessions, therefore, 10 min were allowed for recall without monitoring subjects' progress. Only one subject reported needing (and was given) additional time when queried after 10 min.

**Subjects.** The subjects were 295 UNC-Greensboro introductory psychology volunteers. Each orienting condition included from 26 to 31 randomly assigned subjects (mean = 29), tested in groups of 1 to 30. Booklets from all conditions were distributed randomly in each testing session.

**Results**

Unless otherwise noted, all statistical results discussed hereafter were reliable at or beyond the .05 level. Recall was scored using a gist criterion. Two undergraduates naive with respect to alternative predictions scored each recall protocol. Parts of metaphors (topics or vehicles) recalled alone or mismatched were not scored. Sentences were counted correct only if both scorers agreed, after discussion if necessary, that the original gist was maintained. Although the scorers did not record the frequency with which discussion was necessary, they reported that disagreements were rare after the first few protocols (which were scored in the presence of the first author). The two subsequent experiments substantiated the observation that most reponses in this task are clearly either right or wrong.

Because some subjects failed to rate some metaphors, a proportional recall measure (number recalled/number rated) was computed for each subject. The mean percentages of metaphors recalled in the 10 orienting conditions are listed in Table 1. Individual subjects' recall ranged from 1 to 26 items. The recall data first were analyzed using a one-way ANOVA with orienting condition as the only (between-subjects) factor. As is apparent from Table 1, no differences in recall as a function of condition were found [ $F(9,285) = 1.29$ ,  $MSe = .002$ ]. Therefore,

the effects of orienting condition will not be considered further.

As a first step toward identifying the attributes of metaphors that might affect their recall, the 10 mean ratings obtained for each item were correlated with the (proportional) frequency of recall of each item in each orienting condition. The resulting correlation coefficients are shown in Table 2, where it can be seen that several of the normative dimensions reliably predicted recall in several rating conditions. In half of the conditions (imagery of sentence subject, imagery of sentence predicate, overall imagery, number of interpretations, and felt familiarity), recall was positively and reliably related to all 10 normative dimensions.

To clarify possible relations between recall and the rated attributes of the metaphors, 11 stepwise multiple regression analyses were performed, one for each orienting condition and one for recall summed over the 10 conditions. This procedure provided control over the confounding interrelations among the dimensions reported by Marschark et al. (1983) and allowed evaluation of each dimension's relative contribution in predicting recall. In each analysis, the frequency of recall for each metaphor was used as the criterion variable, and the means of the 10 normative dimensions were entered as predictor variables. Because of the reliable interrelationships among many of the dimensions, both predictor and criterion variables first were converted to z-scores for the purpose of the multiple regression. The 10 regression equations are listed in Table 3. Reading across the table are the criterion variable, the multiple R obtained, and the beta weights of those predictors that had reliable ( $\alpha = .05$ ) and statistically independent relationships to the criterion. Inspection of Table 3 reveals that only the rated number of interpretations available for a metaphor and the rated imageability of its subject (topic) were reliable predictors of recall in all 10 conditions, having been entered first and second into all 10 equations. Those metaphors most frequently recalled generally were those rated as having more interpretations and being higher in subject imagery. The rated semantic relatedness of topics and vehicles, when

**Table 2**  
Intercorrelations among Normative Ratings and Recall in 10 Orienting Conditions of Experiment 1

Normative Dimension	Orienting (Recall) Condition									
	INT	EOI	DOM	GOOD	I	IS	IP	SEM	FAM	COMP
1. INT	.41†	.16	-.11	.12	.21*	.25†	.19*	.04	.23*	.16
2. EOI	.33†	.25†	-.02	.18*	.30†	.40†	.34†	.15	.28†	.19*
3. DOM	.40†	.34†	.04	.25†	.42†	.45†	.38†	.22*	.37†	.33†
4. GOOD	.50†	.37†	.12	.32†	.39†	.42†	.43†	.26†	.39†	.32†
5. I	.43†	.25†	.01	.16	.31†	.36†	.32†	.15	.31†	.20*
6. IS	.34†	.15	-.15	.06	.20*	.31†	.20*	.04	.18*	.13
7. IP	.46†	.22*	-.05	.18*	.28†	.32†	.28†	.13	.31†	.19*
8. SEM	.50†	.36†	.03	.29†	.39†	.46†	.42†	.23*	.39†	.33†
9. FAM	.43†	.21*	.01	.15	.30†	.32†	.26†	.14	.34†	.22*
10. COMP	.40†	.16	-.09	.11	.24†	.29†	.20*	.07	.25†	.14

\*  $p < .05$  †  $p < .01$   $df = 96$

Table 3  
Multiple Correlations Predicting Recall in 10 Orienting Conditions of  
Experiment 1 Using Normative Ratings as Predictors

Orienting Condition	Predictor Variables										
	R	INT	EOI	DOM	GOOD	I	IS	IP	SEM	FAM	COMP
1. INT	.52	.38					.19				
2. EOI	.55	.32					.75				
3. DOM	.63	.28				.70	.34		-.73		
4. GOOD	.61	.44					.35				
5. I	.58	.47			-.35		.52				
6. IS	.55	.26		-.41			.48				
7. IP	.55	.40		-.25			.39				
8. SEM	.67	.44					.65		-.34		
9. FAM	.55	.47			-.32		.46				
10. COMP	.52	.45			-.36		.46				
TOTAL RECALL	.45	.44					.66		-.41		

reliable, was a negative predictor of recall, as were the rated goodness and degree of metaphoricity of the items (see Appendix A).

### Discussion

Mean levels of recall were approximately equal across the 10 orienting conditions of Experiment 1. This finding, however, need not indicate that subjects in all conditions interpreted or encoded the metaphors identically. Indeed, the results of the multiple regression analyses, while generally quite consistent, suggested that different factors affected recall in some conditions. Although the source of this variability is unclear at present, these results do highlight the complexity of memory for metaphor. Further adding to the puzzle, several findings from the present study are clearly at odds with previous findings concerning metaphor comprehension.

First, the consistently reliable effects of topic imagery here contrast with conclusions of Reichmann and Coste (1980) and Honeck (1973) that imagery is relatively unimportant for figurative language comprehension. This finding also is contrary to Paivio's (1979) suggestion that the concreteness of the vehicle should be particularly crucial for comprehension insofar as information-rich images associated with the vehicle guide schematization of the topic. As noted previously, however, the Honeck (1973) and Reichmann and Coste (1980) studies involved proverbs, not metaphors, and Paivio (1979) did not provide any empirical support for this hypothesis. Moreover, the consistent effects of imagery of subject are consistent with previous findings from paired associate tasks, indicating stimulus imagery to be a more potent predictor of memory than response imagery (e.g., Yarmey & O'Neill, 1969).

A second finding inconsistent with previous metaphor research concerns the semantic relatedness of topics and vehicles. Although considerable importance has been laid to semantic relatedness in the metaphor comprehension literature (e.g., Johnson & Malgady, 1979; Malgady & Johnson, 1976), relatedness was not strongly related to recall here, and when it was a predictor, the relationship was negative. Apparently, the semantic overlap between

a topic and a vehicle may be important in arriving at an interpretation of a metaphor, but it is less important for memory and may even interfere if the interpretation is too obvious. The successful integration of several topic-vehicle convergences thus seems more important for memory than their individual retention (Honeck, 1973; Verbrugge & McCarrell, 1977). Consistent with this suggestion, McCabe (1983) found that topic-vehicle similarity was positively related to metaphor goodness only in isolated sentences (in lists). When metaphors were presented in context or in the wording of original speakers, no consistent relationship of relatedness and goodness was observed. Although McCabe's (1983) results suggested that the formulaic nature of the stimuli in many metaphor studies might be responsible, in part, for the relatedness-goodness relationship, that effect was reliable in the Katz et al. (1985) study of poetic metaphors drawn from literary works. Topic-vehicle relatedness thus seems to have some effect on ratings of the goodness of isolated metaphors, but rather less effect on metaphor memory and comprehension (Marschark et al., 1983).

Perhaps the most striking result of this experiment was the consistent prediction of recall, in all 10 orienting conditions, by the rated number of interpretations available for the metaphors. This finding was particularly interesting because the variable has received very little attention in the metaphor comprehension literature, apparently having been considered only in the context of Johnson and Malgady's (1980) perceptual theory of metaphor. Johnson and Malgady proposed that the "best" metaphors are those with only a single meaning. This assumption would suggest that number of interpretations should have been inversely related to metaphoric goodness in Experiment 1. Instead, a positive relationship between these variables was obtained, as was the case in the Marschark et al. (1983) normative study.

The observed effects of number of interpretations reinforce the conclusion that memory for metaphor involves some dimensions not previously identified as important for metaphor comprehension. Depending upon one's assumptions about how a metaphor is encoded, however, the effect of number of interpretations might have two

alternative explanations. Both of these have substantial histories in memory research, and they provide a fortunate context in which to integrate findings concerning memory for metaphors and more general memory theory. According to the general theoretical approach emphasizing the generation of associative networks and spreading activation retrieval mechanisms (e.g., Anderson, 1983; Collins & Loftus, 1975), metaphors with multiple interpretations might be assumed to induce generation of multiple representations (i.e., encoding variability) in the course of comprehension. The potential for reactivation of any one representation at retrieval then would confer an advantage in free recall relative to metaphors with only a single interpretation (Anderson & Bower, 1972; Glenberg, 1979).<sup>1</sup> Alternatively, metaphors rated as having multiple interpretations may have higher probabilities than others of evoking a single, particularly distinctive or elaborated representation ( Craik & Tulving, 1975; Lockhart, Craik, & Jacoby, 1976). Those metaphors may have only one high-probability interpretation, but with several elaborating nuances that give the impression of having multiple meanings in a normative rating task. In this case, encoding a "multiple interpretation" metaphor would involve generation of only one representation, but it would have an advantage in memory because of the enhanced availability of that representation at retrieval relative to metaphors that lack such elaborated, distinctive encodings.

## EXPERIMENT 2

The second experiment was intended to provide a partial replication of Experiment 1 and to examine the two alternative explanations of the number of interpretations effect on free recall. Three orienting tasks were designed so as to differ in the extent to which each would activate alternative interpretations of a subset of the metaphors used in Experiment 1. One condition involved rating the number of interpretations available for each of 36 metaphors; a second required the production of as many interpretations as possible for each metaphor, and a third required production of only a single interpretation for each. If the effects of number of interpretations lie in the generation of multiple encodings for a metaphor, producing multiple interpretations should lead to better recall than producing a single interpretation, and to as good or better recall than rating the number of interpretations. Furthermore, to the extent that writing the alternatives requires more precise and complete encoding than does simply estimating their number, recall in the multiple generation condition should surpass that observed in the rating condition. In contrast, if the effects of number of interpretations lie in the greater likelihood of subjects' storing a single, elaborated interpretation of a metaphor, then drawing their attention to that one interpretation to the exclusion of others should enhance recall (Hunt & Einstein, 1981). Recall following generation of a single good interpretation of the metaphor then would surpass that fol-

lowing either generation of multiple interpretations or estimating the number of alternative interpretations, because either would be distracting.

## Method

**Stimuli.** Thirty-six metaphors were selected from the materials of Experiment 1. Eighteen were relatively high (mean = 2.22) and 18 were relatively low (mean = 1.59) on the rated number of interpretations available for them, the difference being highly reliable [ $t(34) = 9.95$ ]. At the same time, the two material sets did not differ on any other of the nine dimensions rated in Experiment 1, and within each set, the mean ratings of subject and predicate imagery were approximately equal to each other, all  $t(34) < 1.46$ . The 36 test metaphors are listed in Appendix B.

Each stimulus sentence was typed in IBM Courier typeface, photo-reduced and copied onto transparencies that then were made into slides. The 36 test slides were randomly ordered in a single list, preceded and followed by sets of four buffer items also drawn from Experiment 1. The same stimuli were used in all three conditions.

**Design and procedure.** The experiment was described to subjects as a normative study, with no mention of a recall test. The metaphors were presented at a rate of 15 sec/sentence via a carousel slide projector controlled by external timers. During the exposure of each slide, subjects were engaged in one of the three orienting tasks. Those in the RATING condition were read the instructions used in the number of interpretations rating task used in Experiment 1 (see Appendix A). These subjects were told that they were to use the exposure duration of each slide to write their estimate of the number of interpretations possible for each metaphor. Subjects in the MULTIPLE condition were told that sentences would be presented that were noteworthy because each had several alternative interpretations. During the exposure of each metaphor, they were to write as many interpretations as possible. Subjects in the SINGLE condition were told that the stimuli were noteworthy because each had one figurative interpretation besides the (strange) literal one. During the exposure of each slide, they were to discover and write that interpretation. Subjects in all three conditions wrote their answers (ratings or interpretations) in prenumbered test booklets. Following presentation of all of the stimuli, the booklets were collected, and empty booklets were distributed. Subjects then were asked to write as many of the original metaphors as they could recall. Exact wording was requested if possible, but if it was not, subjects were encouraged to write anything else they could remember. Eight min were allotted for recall.<sup>2</sup>

**Subjects.** Fifteen different introductory psychology volunteers served in each of the three conditions. None of the 45 had participated in Experiment 1.

## Results

First, it is noteworthy that the mean number of interpretations provided by subjects in the MULTIPLE condition was reliably lower than that estimated in the RATING condition, 1.52 versus 1.91,  $t(32) = 6.24$ , although the two had the same range (1 to 3) and varied consistently,  $r(32) = .58$ . The possible explanations of this result are several. Subjects in the rating condition may have overestimated, either because they did not distinguish various nuances in meaning from actual differences in interpretation (i.e., they may have provided general feelings without really evaluating the number they generated) or because they felt that higher ratings were "better" than lower ratings (e.g., they might be judged as being more creative). Alternatively, the numbers of interpretations

**Table 4**  
Correlations between 10 Normative Dimensions and Recall in the 3 Orienting Task Conditions of Experiment 2

Normative Dimension	Orienting (Recall) Condition		
	RATING	SINGLE	MULTIPLE
1. INT	.40*	.46*	.44*
2. EOI	.21	.16	.20
3. DOM	-.23	-.30	-.18
4. GOOD	.07	.06	.16
5. I	.16	.08	.23
6. IS	.30	.28	.34
7. IP	.14	.13	.34
8. SEM	-.12	-.22	-.21
9. FAM	.20	.13	.18
10. COMP	.12	.10	.21

\*  $p < .05$

given by subjects in the MULTIPLE condition may be underestimates, either because writing down interpretations interferes with the generation of other alternatives or because the 15-sec delay between items was not long enough to complete the task.<sup>3</sup> Unfortunately, we are not able to rule out any of these possibilities unequivocally with the available data. The low number of interpretations actually provided, however, make it unlikely that subjects did not have sufficient time to generate more alternatives.

A gist criterion similar to that used in Experiment 1 was used in scoring recall. Buffer sentences were not scored, nor were interpretations of the metaphors erroneously recalled as original stimuli (approximately 3% of the total recall). A single undergraduate experimenter scored all of the recall protocols. A random sample of approximately 50% of the protocols also was scored independently by one of the authors. This yielded 95% agreement in "correct" versus "incorrect" judgments of recalled sentences. Almost all of the disagreements involved the same two stimulus sentences, and all were resolved in conference. Subsequently, the remaining protocols were rescored in accord with those resolutions.

The recall data first were analyzed using a 3 (orienting task)  $\times$  2 (material subset-high or low number of interpretations) ANOVA in which orienting task was a between-subjects factor and materials was a within-subjects factor. Surprisingly, no effect of orienting task was observed [ $F(2,42) < 1$ ,  $MSe = 6.30$ ]. The only reliable effect was the superior recall of metaphors rated as having more interpretations as compared with those rated as

having fewer interpretations, in all three conditions (6.7 vs. 4.4, 6.0 vs. 3.9, and 7.1 vs. 3.7, for RATING, SINGLE, and MULTIPLE conditions, respectively) [ $F(1,42) = 78.76$ ,  $MSe = 1.90$ ]. These data were evaluated further by correlating the normative ratings of the metaphors with their respective frequencies of recall. The resulting coefficients are listed in Table 4. Quite simply, rated number of interpretations was the only variable clearly related to recall in any (and all) of the conditions.

To examine the independent effects on recall of the 10 normative dimensions (controlling the effects of their intercorrelations), four stepwise multiple regression analyses were performed, one for total recall and one for each condition separately. Frequency of recall for each metaphor was used as the criterion variable, and the 10 mean normative ratings for each metaphor were the predictor variables. The four resulting regression equations (based on standard scores) are listed in Table 5. In each, number of interpretations was entered first and imagery of subject second as positive predictors of recall. In the SINGLE condition, overall imagery also was a negative predictor of recall when all other variables were controlled. In the MULTIPLE condition, rating comprehensibility also was a positive predictor of recall, whereas the semantic relatedness of topics and vehicles was a negative predictor of recall when all other variables were controlled. In the total recall equation, semantic relatedness was a third reliable (negative) predictor.

Partial recall of metaphors in this experiment was relatively rare, ranging (nonsignificantly) from 4% to 7% in the three conditions [ $F(2,42) = 2.06$ ,  $MSe = .37$ ]. Twenty-one of the 27 items scored as partial recalls actually included both the one- or two-word subject and the predicate noun of the relevant metaphors but were lacking essential modifiers in the predicate. Scoring this way, significantly more partial recalls occurred to multiple interpretation than single interpretation metaphors [ $F(1,42) = 17.08$ ]. Scoring as correct any item that included the correct predicate noun (as did Verbrugge & McCarrell, 1977) had the effect of reducing the already small differences in recall between the orienting conditions and left only six partial recalls: three subjects without predicates and three predicates without subjects. In general, the symmetry of these results supports the notion of wholistic memory for metaphors rather than retention of discrete associations between subjects and predicates (Verbrugge & McCarrell, 1977).

**Table 5**  
Multiple Correlations Predicting Recall in 3 Orienting Conditions of Experiment 2 Using Normative Ratings as Predictors

Orienting Condition	Predictor Variables										
	R	INT	EOI	DOM	GOOD	I	IS	IP	SEM	FAM	COMP
RATING	.54	.45					.36				
SINGLE	.64	.52				-.51	.77				
MULTIPLE	.74	.33					.31		-.75		.64
TOTAL RECALL	.75	.39					.35		-.67		

## Discussion

The results of Experiment 2 replicated the primary results from Experiment 1. Rated number of interpretations and imagery of subjects were again the most potent predictors of recall. These results, together with the inverse relationship between semantic relatedness and recall, provide additional evidence indicating that memory for metaphors depends on different attributes than have been identified previously as most important for metaphor comprehension.

The other major purpose of Experiment 2 was to explore for differences in memory as a function of orienting tasks designed to emphasize multiple or single encodings of metaphors. However, the orienting tasks had no reliable, differential effects on performance. This result has precedents in other studies that have found material effects that were independent of orienting tasks (e.g., Hunt, Elliot, & Spence, 1979). Such results simply indicate that orienting tasks are not always successful in controlling encoding processes and that, in some cases, to-be-remembered material exerts an independent and substantial influence upon encoding. Under these circumstances, the orienting tasks lose their use as analytic tools for inferring the nature of encoding processes. Thus, although the second experiment underscores the importance of number of interpretations in memory for metaphor, no clear conclusion concerning the effects of this variable upon encoding processes is possible.

## EXPERIMENT 3

Given the lack of any effect on free recall as a function of orienting task in Experiment 2, a cued recall task was used in Experiment 3 in a further attempt to clarify the effects on memory of a metaphor's rated number of alternative interpretations. In one condition, sentence subjects (metaphoric topics) served as cues; in the other, sentence predicates (metaphoric vehicles) served as cues. The reasoning relating the cued recall paradigm to number of interpretations appeals to research and theory involving memory for multiple responses that are associated with a single cue. Various manifestations of the interference paradox, notably fan effects (Anderson, 1983) and cue overload (O. Watkins & M. Watkins, 1975), indicate that the more distinct target items associated with a particular cue, the poorer is recall.

On the basis of previous metaphor research, we have assumed here that memory for metaphors involves the encoding of higher order grounds rather than simple associations between subject and predicate terms (Honeck, 1973; Reichmann & Coste, 1980; Verbrugge & McCarrell, 1977). Given this assumption, if multiple-interpretation metaphors are represented as multiple representations in memory, rated number of interpretations might be inversely related to cued recall even though directly related to free recall. Not only would the cue not bear a direct relationship to the abstract memory representa-

tion (regardless of metaphor imageability), but the multiplicity of alternatives interposed between the cue and reintegration of the original metaphor could interfere with retrieval. If multiple-interpretation metaphors are stored in a single, distinctive representation, verbatim recall cues still might not bear a direct relationship to an abstract memory representation, but there would be a one-to-one relationship between cue and target. Cued recall, therefore, would not vary with the number of alternative interpretations, unless a single elaborated representation were more likely to contain topic or vehicle referents, and then cued recall would be positively related to the number of alternative interpretations.

One further prediction in this experiment derives from Paivio's (1979) discussion of the roles of imagery and retrieval in metaphor comprehension. He assumed that the topic and vehicle of a metaphor are encoded in much the same way as stimuli and responses in paired-associate learning. Noting that recall accuracy in the paired-associate task depends particularly on the concreteness of the retrieval cue, he suggested that the concreteness of the metaphoric vehicle should be crucial in metaphor comprehension. This follows from the assumption that the vehicle (which specifies the topic) would promote retrieval of verbal and nonverbal information associated with the topic (see Marschark & Nall, 1985, for further discussion). Consistent with this prediction, Verbrugge and McCarrell (1977) had found metaphoric vehicles to be superior to topics in cueing the remainder of metaphors. Their materials, however, were not controlled for imagery or any other dimensions now known to affect metaphoric processing. The present experiment therefore allowed a better controlled test of Paivio's (1979) position. The prediction from that position would be that rated imagery of predicates should be an important predictor of cued recall, and that predicate (vehicle) cues should be superior to subject (topic) cues.

## Method

**Stimuli, design, and procedure.** The stimuli were the 36 metaphor and 8 buffer slides prepared for Experiment 2. Subjects were engaged in rating the metaphors and then were given an incidental cued recall test. In the orienting task, they rated each metaphor for number of interpretations, writing their answers during the 10-sec intervals between slides. The following cued recall task included two conditions, one in which subjects were given the topics of the 36 test stimuli and asked to provide the vehicles (TOPIC CUE condition), and the other in which they were given the vehicles and asked to provide the topics (VEHICLE CUE condition). All 36 cues for each condition were randomly ordered on two pages of a test booklet. As in the previous experiments, subjects were encouraged to write any sentence parts or ideas if they could not remember exact wording. Ten min were allowed for recall.

**Subjects.** Fifteen subjects served in the TOPIC CUE condition and 15 others in the VEHICLE CUE condition. All 30 were introductory psychology volunteers, none of whom had participated in either preceding experiment.

## Results

Recall was scored as in Experiment 2. Interrater reliability was 96%. The recall data first were analyzed using

a 2 (cueing condition)  $\times$  2 (material subset—high or low number of interpretations) ANOVA. This validated the observation that topics and vehicles were equally effective recall cues (mean recall = 10.8 and 10.1 items, respectively) [ $F(1,28) < 1$ ]. The only reliable effect was the superior recall of metaphors with relatively few rather than many interpretations (means = 11.1 and 9.9 items, respectively) [ $F(1,28) = 6.91$ ,  $MSe = 2.95$ ], contrary to findings from Experiments 1 and 2.

The recall data then were evaluated by computing correlations between the frequency of recall of each metaphor in the separate TOPIC CUE and VEHICLE CUE conditions and the 10 normative dimensions controlled in this experiment. As can be seen in Table 6, rated number of interpretations was not reliably related to recall in either independent group and, consistent with the ANOVA results, the relatively small coefficients obtained were negative. These data were examined further through three stepwise multiple regressions using frequency of metaphor recall in the TOPIC and VEHICLE CUE conditions separately and their sum (total recall) as the criterion variables and mean ratings on the 10 normative dimensions as predictors. Consistent with the previous analyses, the only variable providing significant contribution to any of the three regression equations was the rated imagery of the sentence subjects. This variable yielded multiple  $R_s$  (and beta weights) of .60 overall and .61 and .49 in the separate TOPIC and VEHICLE CUE conditions, respectively.

### Discussion

The third experiment produced several theoretically pertinent results. As in Experiments 1 and 2, recall here was a positive function of topic (subject) imagery. In contrast to those experiments, however, metaphors with fewer interpretations were better remembered here than those with multiple interpretations when recall was cued with metaphoric topics or vehicles. Contrary to predictions from Paivio's (1979) description of the role of imagery in metaphor comprehension, imagery of subject rather than imagery of predicate was a reliable predictor of recall

here, and topics and vehicles did not differ reliably as effective recall cues. The latter result also is consistent with Verbrugge and McCarrell's (1977) suggestion that metaphors are remembered in terms of higher order, emergent relations among explicit and implicit aspects of the topic and vehicle domains, rather than associative relations between the topics and vehicles themselves.

The finding that in cued recall, unlike in free recall, number of interpretations was inversely related to recall can be explained by assuming that with increasing numbers of interpretations for a metaphor, multiple semantic representations become increasingly probable. Apparently, in free recall, where there is no specific information with which to guide retrieval, the more potential pathways there are available, the higher is the probability that one will be found and recall will be successful. In cued recall, however, when a component of a metaphor is activated by presentation of the topic or vehicle, more interpretations cause more interference, reducing the likelihood of recall (Anderson, 1983; Matthews, 1966; Watkins & Watkins, 1975). Such interference is not created when a single, elaborated representation is contacted, but only when distinct representations are accessed by the recall cue.

### GENERAL DISCUSSION

The goals of the present research were to make contact between the study of metaphor and extant ideas about memory and to determine which, if any, of the variables that have been shown to affect metaphor comprehension also affect memory for metaphor. Both of these endeavors were successful to a greater degree than we expected.

As noted earlier, the bulk of recent metaphor research has concerned the comprehension or interpretation of figurative meaning. Diverse investigators have identified the semantic relatedness of topics and vehicles (Malgady & Johnson, 1976, 1980), their imageability (Langer, 1948; Paivio, 1971, 1979; Verbrugge, 1977), metaphor goodness (Johnson & Malgady, 1980), feature saliency (Katz, 1982), and a variety of interpretability measures as playing important roles in metaphor comprehension. In contrast, although several researchers seem to agree on the general nature of memory for metaphor, no specific variables have been investigated in this regard. Discussions of metaphor comprehension by Malgady and Johnson (1976, 1980) and Paivio (1979) suggest the importance of associative relationships between topics and vehicles in the encoding and retrieval of metaphors (see Katz, 1982, for further discussion), but no empirical analyses have been made. Moreover, findings of Verbrugge and McCarrell (1977) and the present study have suggested that the functional memory unit of metaphors might transcend such relations.

The present experiments involved incidental free and cued recall of metaphors following a variety of orienting tasks. In all cases, the rated number of alternative interpretations available for the metaphors and the rated image-

**Table 6**  
Correlations between 10 Normative Dimensions and Recall  
in the TOPIC CUE and VEHICLE CUE  
Conditions of Experiment 3

Normative Dimension	Recall Condition	
	TOPIC CUE	VEHICLE CUE
1. INT	-.15	-.10
2. EOI	.40†	.34*
3. DOM	.15	.03
4. GOOD	-.08	.00
5. I	.49†	.37*
6. IS	.61†	.49†
7. IP	.46†	.41†
8. SEM	.23	.12
9. FAM	.15	.11
10. COMP	.37*	.26

\*  $p < .05$  †  $p < .01$



ability of their topics were the most consistent predictors of recall. Yet, neither of these potent predictors has ever been identified as facilitating metaphor comprehension or interpretation, and none of the variables previously identified as important for metaphor comprehension were found to be positively and reliably related to recall in the present experiments. Semantic relatedness of topics and vehicle, for example, previously has been shown to facilitate comprehension of metaphors, but that variable was negatively related to recall here in all conditions in which it was a reliable predictor (cf. McCabe, 1983).

Rated imagery of subject was a reliable predictor of recall in all 15 independent conditions of the present experiments. Rated imagery of predicate was not a reliable predictor in any of those conditions. These findings, as well as the equality of topics and vehicles as recall cues in Experiment 3, are clearly inconsistent with Paivio's (1979) hypothesis that vehicles should be especially effective as cues for retrieving information in metaphor comprehension. Instead, the topics of metaphoric sentences appear to serve as conceptual pegs for memory much as do stimulus items in paired-associate learning (Paivio, 1971; Yarmey & O'Neill, 1969). This importance of topic imagery in recall is consistent with findings of Marshcark et al. (1983) from their normative, comprehension-rating task and Nall's (1983) study of metaphor comprehension in children. In the latter study, Nall found that topic imagery and whole-metaphor imagery were more important than vehicle imagery in predicting children's metaphor comprehension, as measured by a multiple choice test (see Marschark & Nall, 1985, for complete discussion of the problem area).

Several of the present findings were consistent with the suggestions of Verbrugge and McCarrell (1977), Honeck (1973), and Reichmann and Coste (1980) that memory for metaphor entails the representation of higher order metaphoric grounds rather than the retention of topic-vehicle associations. The positive relationship between free recall and rated number of interpretations in Experiment 1 and 2, however, failed to distinguish whether representation of multiple grounds or of only a single, elaborated ground was responsible for the advantage of multiple-interpretation metaphors.<sup>4</sup> The inverse relationship between cued recall and number of interpretations in Experiment 3 appears to favor the multiple-interpretation hypothesis insofar as previous memory research has indicated that although multiple memories of linguistic information can facilitate free recall (e.g., Glenberg, 1979), the activation of several discrete facts by a single cue can interfere with cued recall (Anderson, 1983; Watkins & Watkins, 1975). That inverse relationship should not have been obtained if multiple-interpretation metaphors were represented as a single, elaborated ground. In fact, greater elaboration (with increasing variations on a core meaning) of a metaphor then would likely facilitate cued recall because there would be greater semantic cohesion between the topic and vehicle (Malgady & Johnson, 1980).

In summary, the two most general findings from this study are that metaphors rated as having multiple interpretations appear to be multiply represented in memory, and that the variables that have been suggested to be most important in memory for metaphors are different from those identified in previous metaphor comprehension research. The latter issue remains open, however, because the previous studies generally have failed to control more than one or two variables at a time. The similarity of the simple correlational results in the present experiments and those obtained in previous comprehension studies nonetheless augurs well for eventual resolution.

## REFERENCES

- ANDERSON, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning & Verbal Behavior*, **22**, 261-295.
- ANDERSON, J. R., & BOWER, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, **79**, 97-123.
- CLARK, H., & CLARK, E. (1977). *Psychology and language*. New York: Harcourt Brace Jovanovich.
- CRAIK, F. I. M., & TULVING, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **104**, 268-294.
- COLLINS, A. M., & LOFTUS, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, **82**, 407-428.
- GILDEA, P., & GLUCKSBERG, S. (1983). On understanding metaphor: The role of context. *Journal of Verbal Learning & Verbal Behavior*, **22**, 577-590.
- GLENBERG, A. M. (1979). Component-levels theory of the effect of spacing of repetitions on recall and recognition. *Memory & Cognition*, **7**, 95-112.
- GLUCKSBERG, S., GILDEA, P., & BOOKIN, H. A. (1982). On understanding nonliteral speech: Can people understand metaphors? *Journal of Verbal Learning & Verbal Behavior*, **21**, 85-98.
- HONECK, R. P. (1973). Interpretive versus structural effects on semantic memory. *Journal of Verbal Learning & Verbal Behavior*, **12**, 448-455.
- HONECK, R. P., & HOFFMAN, R. (Eds.) (1980). *Cognition and figurative language*. Hillsdale, NJ: Erlbaum.
- HUNT, R. R., & EINSTEIN, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning & Verbal Behavior*, **20**, 497-514.
- HUNT, R. R., ELLIOT, J. M., & SPENCE, M. J. (1979). Independent effects of process and structure on encoding. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 339-347.
- JOHNSON, M., & MALGADY, R. (1979). Some cognitive aspects of figurative language. *Journal of Psycholinguistic Research*, **8**, 253-265.
- JOHNSON, M., & MALGADY, R. (1980). Toward a perceptual theory of metaphoric comprehension. In R. P. Honeck & R. Hoffman (Eds.), *Cognition and figurative language*. Hillsdale, NJ: Erlbaum.
- KESS, J. F., & HOPPE, R. A. (1979). Directions in ambiguity theory and research. In G. Prideaux (Ed.), *Perspectives in Experimental Linguistics*. Amsterdam: John Benjamins.
- KATZ, A. N. (1982). Metaphoric relationships: The role of feature saliency. *Journal of Psycholinguistic Research*, **11**, 283-296.
- KATZ, A. N., PAIVIO, A., & MARSCHARK, M. (1985). Poetic comparisons: Psychological dimensions of metaphoric processing. *Journal of Psycholinguistic Research*, **14**, 365-383.
- LANGER, S. (1948). *Philosophy in a new key*. New York: Mentor.
- LOCKHART, R. S., CRAIK, F. I. M., & JACOBY, L. L. (1976). Depth of processing, recognition, and recall. In J. Brown (Ed.), *Recognition and recall*. London: Wiley.
- MALGADY, R. G., & JOHNSON, M. G. (1976). Modifiers in metaphors. *Journal of Psycholinguistic Research*, **5**, 43-52.
- MALGADY, R. G., & JOHNSON, M. G. (1980). Measurement of figura-

- tive language: Semantic feature models of comprehension and appreciation. In R. P. Honeck & R. Hoffman (Eds.), *Cognition and figurative language*. Hillsdale, NJ: Erlbaum.
- MARSCHARK, M., KATZ, A. N., & PAIVIO, A. (1983). Dimensions of metaphor. *Journal of Psycholinguistic Research*, *12*, 17-40.
- MARSCHARK, M., & NALL, L. (1985). Metaphoric competence in cognitive and language development. In H. Reese (Ed.), *Advances in child development and behavior (Vol. 19)*. New York: Academic Press.
- MATTHEWS, W. A. (1966). Continued word associations and free recall. *Quarterly Journal of Experimental Psychology*, *18*, 31-38.
- MCCABE, A. (1983). Concept similarity and the quality of metaphor in isolated sentences versus extended contexts. *Journal of Psycholinguistic Research*, *12*, 41-68.
- NALL, L. (1983). *Dimensions of metaphor comprehension in third, fifth, and seventh graders*. Unpublished master's thesis, Wake Forest University, Winston-Salem, NC.
- ONIFER, W., & SWINNEY, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, *9*, 225-236.
- ORTONY, A. (1975). Why metaphors are necessary and not just nice. *Educational Theory*, *25*, 45-53.
- ORTONY, A. (1979). *Metaphor and thought*. Cambridge: Cambridge University Press.
- PAIVIO, A. (1971). *Imagery and verbal processes*. Hillsdale, NJ: Erlbaum (reprinted).
- PAIVIO, A. (1979). Psychological processes in the comprehension of metaphor. In A. Ortony (Ed.), *Metaphor and thought*. Cambridge: Cambridge University Press.
- PAIVIO, A., & BEGG, I. (1981). *Psychology of language*. Englewood Cliffs, NJ: Prentice-Hall.
- POSTMAN, L., & KNECHT, K. (1983). Encoding variability and retention. *Journal of Verbal Learning & Verbal Behavior*, *22*, 133-152.
- REICHMANN, P., & COSTE, E. (1980). Mental imagery and the comprehension of figurative language: Is there a relationship? In R. P. Honeck & R. Hoffman (Eds.), *Cognition and figurative language*. Hillsdale, NJ: Erlbaum.
- TOURANGEAU, R., & STERNBERG, R. J. (1981). Aptness in metaphor. *Cognitive Psychology*, *13*, 27-55.
- VERBRUGGE, R. (1977). Resemblances in language and perception. In R. E. Shaw & J. D. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology*. Hillsdale, NJ: Erlbaum.
- VERBRUGGE, R., & MCCARRELL, N. (1977). Metaphoric comprehension: Studies in reminding and resembling. *Cognitive Psychology*, *9*, 494-533.
- WATKINS, O., & WATKINS, M. (1975). Buildup of proactive inhibition as a cue overload effect. *Journal of Experimental Psychology: Human Learning & Memory*, *82*, 261-275.
- YARMEY, A. D., & O'NEILL, B. J. (1969). S-R and R-S paired-associate learning as a function of concreteness, imagery, specificity, and association value. *Journal of Psychology*, *71*, 95-109.
- YOUNG, D. R., & BELLEZZA, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *8*, 545-559.

## NOTES

1. With regard to metaphors, the encoding variability alternative is not compromised by the findings of Postman and Knecht (1983) and Young and Bellezza (1982) indicating better memory in encoding constancy than encoding variability situations. Those studies demonstrated that interference can be created when the same target items are embedded in different verbal contexts on different presentations. Such findings are not directly relevant to the variable encoding assumed to occur here as different interpretations of the same sentences are considered in a single presentation.

2. Due to experimenter error, one metaphor low on the number of interpretations dimension was omitted from the stimulus presentation. The most closely matched metaphor high on that dimension therefore was excluded from scoring and analysis. Statistical analyses reported therefore are based on 34 stimuli. Experiment 3 included all 36 test items.

3. We appreciate Neal Kroll's clearly distinguishing these alternatives.
4. This issue corresponds to a similar empirical question with regard to the number of readings generated in comprehension of ambiguous statements (Kess & Hoppe, 1979; Onifer & Swinney, 1981). Indeed, metaphors might be considered one type of ambiguous sentence, having at least one literal (false) reading and at least one figurative reading.

## APPENDIX A

### Rating Instructions Used in Experiment 1

#### Number of Alternatives (Complete)

Consider the following sentence: "When the wind blew the house shook." You probably understand or comprehend what that sentence means. Some sentences, however, often appear to be nonsense or take a long time to understand. Consider, for instance, the following sentence: "When the show was over, the gruesome split." This sentence is probably nonsense to you, although with effort you could probably make some sense out of it. For example, if you thought of The Gruesome as the name of a travelling theatrical group, the sentence could be quite comprehensible.

Some sentences can be given more than one interpretation. Consider some possible interpretations of the sentence: "Truth is a sword." Some possible interpretations could be: 1) truth can be a weapon, 2) truth is long and straight, 3) truth cuts through lies. You might be able to think of other interpretations.

In this experiment, you will be presented with a list of sentences and will be asked to decide on the number of different interpretations that you can think of for each. If you can't think of any, please put a 0 (zero) in the space provided at the end of the sentence; if you think of one, write 1 in the space, and so on. Don't be concerned about the number of times you use a particular number, as long as each is your true judgment.

Below are several examples. Indicate the number of different alternative interpretations you can give to each sentence by writing the appropriate number in the space provided. . . .

(Excerpts of remaining instruction sets)

#### Semantic Relatedness

Consider the following two sentences: 1) Cities are beehives; 2) The city is a library of lights. In each of these sentences, two objects are compared and related to each other. . . . Your task will be to decide how closely related in meaning are the two objects of each sentence. . . . In a sentence like "A city is a large town," for example, CITY and LARGE TOWN probably would be considered fairly close in meaning. However, in a sentence like "A city is a swamp of individuals," CITY and SWAMP OF INDIVIDUALS would probably be considered to have very little in common. Your ratings should be made on a seven-point scale where 1 is the LOW SIMILARITY end and 7 is the HIGH SIMILARITY end. . . .

#### Imageability

Phrases and sentences are known to differ in their ability to arouse mental images. Some sentences arouse mental images (that is, sensory experiences such as mental pictures) very quickly and easily whereas other sentences may only do so with difficulty or not at all. . . . The purpose of this experiment is to determine the ease with which mental images are formed for different sentences . . . you are to rate . . . the ease or difficulty with which they arouse mental images. If you find it easy to form an image to a particular sentence, then you should give

it a HIGH IMAGERY rating. If, however, you find it difficult to form an image to the sentence, then it should be given a LOW IMAGERY rating. . . .

### Imageability of Sentence Subjects

Phrases and sentences are known to differ in their ability to arouse mental images. Some sentences arouse mental images (that is, sensory experiences such as mental pictures) very quickly and easily whereas other sentences may only do so with difficulty or not at all. Also, parts of sentences differ in the ease with which they arouse imagery. Sentences can be considered to consist of two parts: the SUBJECT (i.e., the topic of the sentence, often the "actor") and the PREDICATE (i.e., what the topic is relating to or what the actor is doing). In the sentences that follow, the SUBJECT will always come at the beginning of each. . . . The purpose of this experiment is to determine the ease with which mental images are formed for the SUBJECTS of the sentences on the following pages. . . .

### Imageability of Sentence Predicates

Phrases and sentences are known to differ in their ability to arouse mental images. . . . In the sentences that follow, the PREDICATE will always come at the end of each. . . . The purpose of this experiment is to determine the ease with which mental images are formed for the PREDICATES of the sentences on the following pages. . . .

### Comprehensibility

Consider the following sentence: "When the wind blew the house shook." You probably understand or comprehend what that sentence means. Some sentences, however, often appear to be nonsense or take a long time to understand. Consider . . . "When the show was over, the gruesome split." This sentence is probably nonsense to you, although with effort you could probably make some sense out of it. For example, if you thought of The Gruesome as the name of a travelling theatrical group, the sentence could be quite comprehensible. . . . You will be asked to determine how easy or difficult each one is to understand. If you find a sentence very easy to make sense of, or comprehend, you should give it an EASY sense rating. If, however, you find it difficult to make sense of, or comprehend, a sentence you should give it a DIFFICULT sense rating. Your ratings should be made on a seven-point scale where 1 is the end of the scale denoting difficulty-in-comprehending and 7 is the end of the scale denoting ease-in-comprehending. . . .

### Metaphor Goodness

This study is concerned with the subjective goodness of metaphors. A metaphor is a type of sentence in which one object is compared to another in a non-literal way. For example, consider the following sentence: "An airplane is a motorized glider." This sentence relates airplanes and gliders. It is *not* a very good metaphor because the relationship is one that is fairly literal, since an airplane *is* essentially a glider with engines. Contrast . . . "An airplane is a migrating bird." In this case, airplanes are related to birds. The sentence *is* a metaphor because, literally speaking, an airplane is not a bird. Nonetheless, you probably can interpret the sentence and get its meaning: an airplane is like a bird because it can fly. . . . In this experiment . . . we are interested in determining how GOOD you think each metaphor is. That is, some metaphors make a comparison in a particularly apt and pleasing way. Others are poor figures of

speech and poor descriptions. . . . rate each of the following metaphors on how GOOD you think they are as figures of speech . . . ["very poor metaphor" to "very good metaphor"]

### Ease of Interpretation

This experiment concerns the ease or difficulty of understanding metaphors. . . . For example, consider the following metaphoric sentence: "An airplane is a bird." This sentence relates airplanes and birds. It is a metaphor because the relationship is one that is not literal: an airplane *is not* a bird. The sentence is easy to interpret, however, because both airplanes and birds fly. . . . Contrast . . . "An airplane is a suitcase." In this case, airplanes are related to suitcases. It still may be a metaphor because, literally speaking, an airplane is not a suitcase. Nonetheless, you probably would have trouble interpreting. . . . However, an airplane may be like a suitcase in that things are packed into both, both move over long distances, and so on. Naturally, some sentences comparing two things are nonsensical, or hard to interpret, whereas in others you can see the point very easily. . . . We are interested in determining how easily you are able to interpret each metaphor. Those sentences that you are able to interpret very easily should be given high ratings; those that you can interpret only with difficulty or not at all should be given low ratings . . . ["difficult to interpret" to "easy to interpret"].

### Metaphor: Felt Familiarity

Phrases and sentences are known to differ in the frequency or familiarity of the ideas they express. Consider the following sentence: "My girlfriend is a flower." The idea expressed, that the girlfriend has qualities (beauty, fragility) suggested by flowers has probably been experienced by you to some degree, perhaps frequently. The ideas expressed by other sentences will have been experienced by you even more frequently, and hence will be more familiar. . . . The ideas of other sentences will have been experienced only infrequently, if at all. The purpose of this experiment is to determine *your* familiarity with the ideas expressed in different sentences. . . . If you are very familiar with the idea of a particular sentence, you should give it a relatively high FAMILIARITY rating at or near 7. If you are unfamiliar with the idea of a particular sentence, you should give it a low rating around 1. . . .

### Degree of Metaphoricity

Some sentences can be interpreted literally. For instance, sentences like "Lassie is a dog," or "My girlfriend is a female." Other sentences are not meant to be interpreted literally, but nonetheless make sense on a figurative level. For example, consider the sentence "My girlfriend is a rose." On a literal level, the sentence can be considered nonsense: A woman is not a flower. On a figurative level, however, the sentence can be considered to make sense. Presumably, a person who compares his girlfriend to a rose is making a point that both she and a rose share some elements in common, from his perspective. . . . In a sentence like "My boyfriend is an animal," the intention may be either literal or figurative. . . . you will have to decide, for each, the DEGREE to which it is literal or metaphorical. If you believe a sentence to be HIGHLY FIGURATIVE in its intent, you should give it a rating at the figurative end of the scale. If you believe that a sentence is HIGHLY LITERAL, it should be given a rating at the literal end of the scale. . . . Your ratings should be made on a seven-point scale, where 1 is the figurative end and 7 is the literal end. . . .

**APPENDIX B****Stimulus Metaphors Used in Experiments 2 and 3****High Number of Interpretations**

1. Love is a flower. (3.07)
2. Greed is a hungry buzzard. (2.00)
3. A scientist is a spawning trout. (2.00)
4. A beautiful woman is a desert flower. (2.25)
5. The mind is a sponge. (2.31)
6. An old man is a living history book. (2.11)
7. A friend is a ray of sunshine. (2.43)
8. Poems are the seeds of culture. (2.20)
9. Books are treasure chests of information. (2.30)
10. A rumor is a plague.\* (2.00)
11. The bible is the cement of the church. (2.04)
12. Loneliness is a desert. (2.26)
13. Divorce is the earthquake of the family. (2.00)
14. Babies are angels. (2.31)
15. Old schoolteachers are encyclopedias. (2.34)
16. Nature is a vast laboratory. (2.14)
17. Scientific research is mountain climbing. (2.08)
18. Respect is a precious gem. (2.15)

**Low Number of Interpretations**

1. A white rabbit's fur in winter is a ready-made suit of long underwear. (1.46)

2. A butterfly is a winged rainbow. (1.54)
3. An exaggeration is an invitation to falsehood. (1.52)
4. Smoke is a fire's calling card. (1.50)
5. An umbrella is a portable roof. (1.54)
6. A fan is a private summer's breeze. (1.62)
7. A rocket is a bullet to the stars. (1.50)
8. Chessmasters are computers.\* (1.58)
9. Genes are blueprints. (1.62)
10. Work is the opium of successful people. (1.67)
11. The mosquito is a vampire. (1.69)
12. Dictators are the stranglers of liberty. (1.52)
13. Indecision is a whirlpool. (1.79)
14. Tree trunks are straws for thirsty leaves and branches. (1.69)
15. Craters are the moon's dimples. (1.48)
16. Terrorists are hawks among the fowl. (1.63)
17. Thunderclouds are draperies pulled across the sun. (1.48)
18. Freedom is a breath of fresh air. (1.75)

\*Omitted in Experiment 2

*Note—Mean numbers of interpretations noted in parentheses (see Experiment 1).*

(Manuscript received August 13, 1984;  
revision accepted for publication April 17, 1985.)