Metaphoric Comprehension: Studies in Reminding and Resembling

Robert R. Verbrugge† and Nancy S. McCarrell††

ABSTRACT

The theoretical problems posed by metaphoric comprehension are discussed in the context of experiments on prompted recall. Listeners heard sentences of the form "Topic is (like) Vehicle." In most cases, a statement of the implicit resemblance (the "ground") was very effective in prompting recall of its related metaphor. This result could not be attributed to the activation, transfer, or additive combination of pre-existing properties of the topic and vehicle terms or to pre-existing associations between grounds and sentence terms. It is argued that the vehicle domain guides a novel schematization of the topic domain, that the perceived resemblance is a higher-order relation among entities (both explicit and implicit) in each domain, and that this abstract relation constitutes the "functional memory unit." Prompted recall may begin with recognition of this previously experienced relation.

†Also University of Connecticut, Storrs.

††George Peabody College for Teachers, Nashville, Tennessee.

Acknowledgment: This research was supported in part by a fellowship to Robert Verbrugge from the University of Michigan Society of Fellows; by grants to the University of Minnesota, Center for Research in Human Learning, from the National Science Foundation (GB-35703X), the National Institute of Child Health and Human Development (HD-01136), and the Graduate School of the University of Minnesota; and by a grant (HD-00973) from the National Institute of Child Health and Human Development to George Peabody College for Teachers. We wish to thank John Bransford, James Jenkins, Dane Harwood, James Reaves, Robert Shaw, and Buford Wilson for their advice and encouragement throughout this research and Mark Jaffee for his assistance in data analysis. A preliminary report on Experiments I and II was made to the Midwestern Psychological Association, Chicago, May 1973. The research was undertaken while both authors were at the Center for Research in Human Learning (University of Minnesota). The Human Performance Center (University of Michigan) and Haskins Laboratories provided assistance during preparation of the manuscript and this support is gratefully acknowledged.

Metaphoric language invites a "perception of resemblances,"¹ and the invitations come in many forms. Examples of metaphors (strictly defined) are these: My lawnmower is a wild animal; The children galloped to the cafeteria; Billboards are warts on the landscape. In these cases, a resemblance is communicated by forms that assert or presuppose an identity. Similes and analogies are less bold since they directly assert a relation of similarity: The freeway is like a snake; He runs as fast as a cheetah. Beyond these, there are dozens more "hedge" forms by which similarities are expressed (Lakoff, 1972); for example, George resembles a truck driver; Judy is kind of a prima donna. In each of these metaphoric forms, two domains are being compared: a topic (traditionally called the tenor; Richards, 1936) and a vehicle (that to which the topic is being compared). The topic is usually mentioned explicitly, but in such forms as proverbs, parables, and allegories it must be supplied by the comprehender. Similarly, the vehicle may be mentioned explicitly or it may simply be alluded to, as in the galloping sentence above. The resemblance between the topic and vehicle domains is traditionally called the ground (or tertium comparationis). The ground is occasionally made explicit (as in the cheetah sentence above), but usually it is the reader's or listener's task to discern the resemblance. The task for psychologists, in turn, is to characterize the structure of the apprehended resemblance, its relationship to the terms that appear in a sentence, and the process by which the resemblance is discerned.

Psychologists and linguists have devoted comparatively little attention to the meaning and comprehension of metaphoric language. Part of the explanation may lie in the long tradition in epistemology and rhetoric that stresses the categorization of reality in terms of elementary sensory or semantic features, the sharply defined and enduring character of these features, and the relative stability of their interrelations. If such a semantics is presupposed, metaphor can pose a special problem for explanation, since it often demands that we accept a categorization radically different from what is familiar or conventional. It is a short step to viewing metaphor as an illogical and even freakish language form—an object of universal fascination, perhaps, but one that resides at the periphery of ordinary language use. This academic attitude ignores what seems obvious to casual observation: metaphoric language is endemic to ordinary communication. It is common in day-to-day conversation, narrative, popular songs, newspaper articles, effective teaching, and problem solving. In fact, metaphor may be basic to all growth in understanding, whether in the playroom, the classroom, the psychotherapeutic setting, the scientific laboratory, or the theater (see Hesse, 1966; Langer, 1967; Verbrugge, 1977; Pollio, Barlow, Fine and Pollio, 1977).

Since appreciation of the importance of metaphor has developed only recently in psychology, research on metaphoric comprehension (particularly in adults) has been sparse. Though the research is difficult to classify

¹We have borrowed this phrase from Aristotle, whose views on poetic language are expressed in his Poetics and Rhetoric. A summary may be found in Hawkes (1972).
systematically, it is convenient to identify two traditions: associationism and transformational linguistics.

Associationism proposes that words are associated with an array of elemental ideas, concepts, images, and combinations thereof, and that a probability or strength can be assigned to each of these links. Sentence meaning is some kind of composite of the associations to constituent elements. Metaphors are viewed as fortuitous, low-probability associations, governed by the usual laws of conditioning and transfer. One option is to view the topic and vehicle as having common associates: words with "stimulus equivalence" are linked when producing the metaphor, and comprehension involves activating the common associate (see Asch, 1955; Skinner, 1957). A related view is that metaphors involve the substitution of a response for one that is more typical and appropriate (Osgood, Suci and Tannenbaum, 1957; Brown, Leiter and Hildum, 1957; Koen, 1965). For example, The baritone's voice was heavy might be spoken in response to hearing a singer's voice, due to the strong associations between low-pitched voice, large body, heavy, loud, etc., in prior experience. Comprehension involves activating these high-frequency ("literal") associates and linking them to the topic. While the more sophisticated theories of associative networks (for example, Anderson and Bower, 1973; Norman and Rumelhart, 1975) have seldom been applied to metaphoric sentences, they propose representational structures and procedures that are similar to those just described. Comprehension of a metaphoric sentence would presumably involve detecting common associated predicates in the network or transferring predicates from one node to another.

A second influential approach to the psychology of metaphor is an outgrowth of transformational linguistics. In the semantic systems proposed by Katz and Fodor (1963) and Chomsky (1965), sentence constituents were indexed in a lexicon by grammatical category, a set of distinctive semantic features, and selection restrictions that defined the contexts in which a term could appear. Expressions that failed to honor these restrictions were labeled semantically unacceptable, anomalous, and deviant. Among this riff-raff of rejected word strings were many varieties of figurative language, including metaphor. Other linguists, not wishing to lose metaphor as an object of linguistic description, have suggested that special rules be added to a grammar to permit interpretation of these "deviant" sentence forms (for example, Weinreich, 1966; Bickerton, 1969; Leech, 1969; Matthews, 1971). Common strategies have been to suspend selection restrictions temporarily, to ignore incompatible feature values, or to alter the standard feature descriptions for terms (for example, by reassigning values to some of their features). These are temporary alterations to the language device, allowing it to process abnormal inputs that would otherwise bring it to a grinding halt. To a large extent, these efforts have shared the basic assumptions of the Katz and Fodor (1963) model: metaphor is a semantic violation; its identity and interpretation are to be characterized without reference to the intentionality, nonlinguistic knowledge, or processing strategies of language users; and the special rules operate on stable semantic feature descriptions associated with terms. On the latter point, this approach is similar to traditional associationism, except that a highly constrained structural organization of features is proposed.
The linguistic approach to metaphor sharply distinguishes between sentences that are well formed and anomalous, normal and deviant, acceptable and unacceptable. Many psychologists of language have accepted this dichotomy, focusing their research on "rule-governed" language and contrasting its processing with that of "anomalous" language (for example, Marks and Miller, 1964; Steinberg, 1970; Epstein, 1972; Collins and Quillian, 1972; Smith, Shoben and Rips, 1974). In the few cases in which metaphoric "anomaly" has been the focus of psychological research, the characterization of meaning is similar to that found in associative accounts. For example, Johnson, Malgady and Anderson and Malgady and Johnson (1976) have attempted to define the operations on two partially incompatible feature sets that could yield the appropriate ground (that is, common associated features) as a product. Kintsch (1972, 1974) has argued that metaphors are anomalous surface forms produced by condensation of deep-structure assertions of similarity. In this model, certain "lexical implications" and properties are already associated with both the topic and the vehicle, and comprehension includes a search for associations shared by the two terms.

It is important to determine why associative and linguistic models have shown only localized and transitory success as theories of metaphoric language. We believe that two important hindrances to success have been the following.

(1) Metaphor has been treated as uniquely ambiguous, imprecise, and illogical. In most logics of this century (including that underlying semantic feature theory), meaning is assumed to be sharply bounded, that is, the criteria for ostensive application of a term to a referent are (in principle) precisely and unambiguously defined. Imprecision in language use is attributed to the difficulty encountered by a speaker-hearer in relating the criteria to a specific situation, that is, it is a "performance" phenomenon. Verbal ambiguity, therefore, could result from poor viewing conditions, inattention, carelessness, immaturity, or psychopathology. If metaphor is viewed as an imprecise application of terms to referents, it is a short step to interpreting the metaphoric productions of adults, children, schizophrenics, and poets as "deviant." But more important than this invidious labeling is the conclusion drawn about comprehension: to understand the anomaly one must rationalize it according to the sharply defined constraints that apply to ordinary language. Accordingly, most of the recent accounts of comprehension assume that the listener must "normalize" a metaphor, that is, intuit the literal (precise) meaning that must have been intended.

Dissatisfaction with this view of meaning criteria has grown in recent years. An increasing number of linguists and psychologists have come to believe that semantic feature classification is inadequate for explaining the flexibility and precision of ordinary language (for example, Bolinger, 1965; Cohen and Margalit, 1972; Rosch, 1973; Anderson and Ortony, 1975; Bransford,

---

McCarrell and Nitsch, 1976). One leitmotiv in this dissent is the belief that the underlying criteria for word use are not sharply defined; they are "fuzzy" and abstract constraints. One goal of current theoretical efforts is to understand how precision may be achieved by the application of the constraints in particular contexts (see Lakoff, 1972; Bransford and McCarrell, 1974). If the standard uses of terms are only fuzzily bounded, the distinction between metaphoric and literal language itself becomes fuzzy, and the goal of rationalizing one in terms of the other becomes suspect. In a fuzzy logic, the use of a term is always metaphorical in the following sense: a new context of use has only a sufficient resemblance to prior contexts of use. If we say This penguin is a bird or This creature is a penguin, we are making an assertion about a sufficient resemblance to prototypical constraints on birdiness or penguinicity. The process is very similar when we say My daughter is a bird or That cloud is a penguin; again, these are motivated by the applicability of a set of abstract constraints to a novel instance. Thus, the apparent precision and primacy of literal language dissolves when we realize that all language use occurs in novel contexts, and that these contexts are related by a sufficient resemblance, not an identity defined by invariant criterial features. Metaphoric and literal assertions seem to part company over how exhaustively the conventional constraints apply, not in precision. (Compare a penguin-shaped cloud, a portly gentleman in a tuxedo, and a real penguin.)

(2) A second major hindrance to success in developing a theory of metaphor has been the characterization of grounds in terms of common features and common associations. Metaphoric comprehension has been treated as a kind of concept formation task in which the concepts are "attributive," that is, word meaning is defined by a set of associated attributes. The process is one of "subtractive" concept formation, since shared attributes become part of the ground, while conflicting attributes are ignored. The attributes (features, properties) are treated as substantive building blocks of identity, both in the narrow sense of linguistic meaning (they are elements that concatenate to form word meaning) and the broader sense of knowledge about the referent (they are elements that concatenate to form factual knowledge). The underlying theoretical metaphor has changed little through the long history of associationism: attributes are substantive atoms.

We need to consider carefully whether attributive concepts are sufficient to characterize the grounds of metaphors. Many metaphors draw attention to common systems of relationships or common transformations, in which the identity of the participants is secondary. For example, consider the sentences: A car is like an animal, Tree trunks are straws for thirsty leaves and branches. The first sentence directs attention to systems of relationships among energy consumption, respiration, self-induced motion, sensory systems, and, possibly, a homunculus. In the second sentence, the resemblance is a more constrained type of transformation: suction of fluid through a vertically oriented cylindrical space from a source of fluid to a destination. In each case, the substantive components of the two domains show little or no resemblance. Translating the relationships into attribute lists is an awkward and unbounded process and may be impossible in principle.
There have been many efforts to characterize such systems of relationships or "schemata," to distinguish them from attributive concepts, and to argue against the adequacy of attributive concepts as the primary basis for conceptual knowledge (for example, Cassirer, 1923; Piaget, 1950; Jenkins, 1966; Bransford and Franks, 1973; Weimer, 1973). For present purposes, we will speak of these relational systems as abstract relations, to emphasize that the structure of resemblance is primarily abstract.

A particularly useful characterization of such relations is found in the discussion of event perception by Shaw, McIntyre and Mace (1974). These authors characterize an event in terms of a transformational invariant (a kind of transformation exerted over a structure, for example, rotation) and a structural invariant (what the transformation leaves invariant, for example, spherical shape). Either type of invariant or both can serve as the basis for a resemblance. For example, in the tree trunk sentence, the flow of fluid is a transformational resemblance: the transformation leaves the tubular structure and the volume of fluid invariant in each domain. Since both the tree trunk and the straw have a tubular structure, this constitutes a structural resemblance that enhances the strength of the metaphor. It is tempting to view the structural resemblances as attributes of the traditional kind. It is important to keep in mind, however, that such invariants always presuppose some transformation or system of relationships, and that these are contextually variant. Thus, in Tree trunks are pillars for a roof of leaves and branches, the structural invariant is a solid column rather than a hollow tube. The tree trunk is not the same "structure" in each case; for this reason, a fixed set of properties could not characterize its role in the two different metaphors. In general, attributive concepts fail by overlooking transformational resemblances, by assuming that the resemblances draw on a fixed, contextually invariant set of structural primitives, and by assuming that structural primitives are substantive in kind (rather than abstract or mathematical).

The research reported here focuses on the structure of metaphoric resemblances. Identifying the structure of grounds is a crucial prerequisite to studying how they are discerned. Their structure places important constraints (and demands) on the class of process models one might consider. Traditional definitions of the ground in terms of shared attributes led naturally to models involving feature search, comparison, weighting, and transfer. It is important to determine whether features associated with the nominal terms (objects) in a metaphor are an adequate basis for defining the resemblance discovered by the ordinary listener. The event or relationship in which the objects participate may be more critical in defining the resemblance. If so, a different class of comprehension models is necessitated, in which, for example, salient transformations over the vehicle domain are applied over the topic domain.

The accessibility of acquisition material to recall can provide a sensitive symptom of how the material was interpreted. It is becoming increasingly clear that a person's "orienting task" (whether adopted autonomously or at the experimenter's request) has as distinctive an effect on recall as the properties of the materials themselves (see Jenkins, 1974; Craik and Tulving, 1975). Prompted recall is especially useful as a measure of
comprehension, since it is differentially sensitive to components that are central to sentence meaning (Blumenthal, 1967; Blumenthal and Boakes, 1967; Perfetti and Goldman, 1974), and it is sensitive to information supplied implicitly by the comprehender (Tulving and Thomson, 1973; Barclay, Bransford, Franks, McCarrell and Nitsch, 1974; Anderson and Ortony, 1975).

In the case of metaphoric sentences, prompted recall may provide a sensitive measure of the presence of inferential activity during comprehension, the kind of resemblances inferred, and the context specificity of a topic’s interpretation in different metaphors. Specifically: (a) If an abstract relation is central to what is comprehended from a metaphor, a verbal precis of the relation should be an effective prompt for the sentence's recall (even if no terms in the precis match terms in the original sentence). Abstract resemblances of this sort have proven to be effective prompts for recall of proverbs (Bühler, 1908; Honeck, Reichmann and Hoffman, 1975). (b) If the topic is interpreted uniquely in different metaphors (for example, as a participant in different types of events), then a possible "ground" should only prompt recall of the topic when it specifies the relevant type of event or relationship. For example, the ground for the tree trunks-straws metaphor might be summarized verbally as follows: are tubes which conduct water to where it's needed. This phrase might effectively prompt recall when tree trunks have been compared to straws, but it may not be effective when tree trunks have been compared to pillars, even though it expresses a perfectly valid property of tree trunks. A more effective prompt for the tree trunks-pillars metaphor might be provide support for something above them, since it expresses the resemblance which is specific to the pillars context of interpretation. By using pairs of acquisition lists with common topics and prompting recall with possible grounds, one can test whether such specific interpretations are made. Previous studies on prompted recall have demonstrated this kind of "encoding specificity" for terms in literal sentences and word lists (for example, Thomson and Tulving, 1970; Anderson and Ortony, 1975).

In the experiments reported here, we have studied metaphors that are expressed linguistically, explicitly, and in sentence form, that is, cases where a perceived resemblance is communicated through words, where both the topic and the vehicle are explicitly mentioned, and where the comparison is made within a single sentence rather than in a text or a discourse of greater length. We have used two sentence forms, metaphor ("A is/are B") and simile ("A is/are like B"), and the grounds are combinations of both transformational and structural resemblances. Hypotheses based on abstract relations will be tested in parallel with a series of recall models framed in the language of features. The effort throughout this study is to identify the structure of the comprehended resemblance and its relationship to the terms in a metaphoric sentence.

**EXPERIMENT I**

This experiment tested whether the ground of a metaphor can be an effective prompt for its recall. The design of the study crossed two acquisition lists (with matched sets of topics) with two sets of recall prompts. Subjects received ground prompts that were all relevant or all
irrelevant to the original list of metaphoric sentences.

The rationale for this design was as follows: if a verbal statement of the ground successfully prompts subjects' recall, one might challenge the conclusion that the ground had been inferred during an acquisition process guided by the vehicle. Since the ground states a property that is true of the topic, it might serve as an effective prompt irrelevant of any special interpretation guided by the vehicle. Semantic network and semantic feature theories both suggest that major constituents of a sentence independently activate an array of associated predicates or attributes. Thus, the ground in question may be activated whenever the topic appears. For example, both are tubes which conduct water to where it's needed and provide support for something above them may be activated in response to either acquisition sentence about tree trunks and, therefore, might appear in the record of either event. Alternatively, acquisition sentences could be stored more-or-less verbatim, and the subject's strategy at recall could be to scan this record for a topic that contains the ground in its feature list or for which a path to the ground can be found in the network.

To control for these possibilities, two kinds of prompts may be used: (a) a set of "relevant grounds" in which each prompt is relevant to the sense of an acquisition metaphor, or (b) a set of "irrelevant grounds" in which each prompt is irrelevant to the sense of a particular metaphor, but is nonetheless true of its topic. If the vehicle does affect interpretation of the topic, relevant grounds should be more effective as prompts than irrelevant grounds. To insure that this difference between the two sets of grounds is not artifactual, one can present a second group of subjects with a list of acquisition metaphors (using the same topics) for which the formerly "relevant" grounds are now irrelevant and the formerly "irrelevant" grounds are now relevant. The ordering of prompt effectiveness should reverse, even though the topics involved are the same in both cases.

Method

Materials. Two lists of 14 metaphoric sentences were prepared (lists A and B). The topics in each list were the same, while the vehicles were different. For example, tree trunks were compared to pillars in List A and to straws in List B. The various topics and vehicles were kept as distinct as possible; with the exception of the paired topics, no nouns or close synonyms were repeated elsewhere in either list. This was intended to minimize systematic errors in recall. The lists were recorded on audio tape by an adult male speaker using a natural speaking pace, amplitude, and intonation contour. Each sentence was spoken twice. There was a 5-sec pause between the repetition and the next sentence in the list. Topics appeared in the same order in each list.

A "ground" was prepared for each of the 28 metaphoric sentences for use as a prompt. The ground took the form of a predicate expression. It was intended to summarize the major resemblance underlying the metaphor, but was not assumed to be an exhaustive interpretation. The following are further examples of acquisition sentences (and grounds) used in the study.
Skyscrapers are honeycombs of glass. (are partitioned into hundreds of small units)

Skyscrapers are the giraffes of a city. (are very tall compared to surrounding things)

Billboards are warts on the landscape. (are ugly protrusions on a surface)

Billboards are the yellow pages of a highway. (tell you where to find businesses in the area).

The acquisition sentences were written to keep the 28 grounds as dissimilar as possible, again to avoid systematic intrusions in recall. In particular, the pairs of grounds for each topic were chosen to be as unrelated to each other as possible. Each ground avoided content words appearing in the related sentences and terms that are typically constrained to either the topic or vehicle context.

The grounds were assembled into two sets of prompts, grounds A and B, for use in recall. Grounds A were the 14 grounds relevant to the sentences in List A; Grounds B were relevant to List B. The grounds were typed on individual slips of paper, with ample space for subjects to write out a sentence during recall. Each set of prompts was presented in booklet form; a blank slip of paper on top of the booklet obscured the first prompt from view.

In addition, prompt booklets containing the topics and vehicles from the original sentences were prepared. Topics A and B were identical and contained the full-subject noun phrases from the 14 sentence pairs. Vehicles A and Vehicles B contained the vehicles from the related acquisition lists. In some cases the full predicate noun phrase was not included. If a word or phrase in the predicate (for example, leaves and branches) was related to the topic domain (tree trunk), it was excluded from the vehicle prompt.

The order of prompts in all booklets was randomized with respect to the acquisition order, and the same order was used in all cases (that is, the order of correct recall would be the same).

Subjects. Subjects were 96 undergraduates enrolled in an introductory psychology course at the University of Minnesota. They received extra credit for their participation. Subjects were randomly assigned to one of two list conditions, List A or List B. In each condition, 8 subjects received the related topic prompts (Topics A or B), 8 received the related vehicles (Vehicles A or B), 16 received Grounds A, and 16 received Grounds B.

Procedure. In each session a group of subjects sat in a small experimental room facing a tape recorder placed on a desk at the front. The experimenter informed them that they would hear a series of metaphoric sentences describing various types of people, emotions, objects, and so on. They were asked to listen to the sentences and think about what each one was trying to express. No mention was made of a subsequent recall task. After playing List A or B, the experimenter informed subjects that they would receive a booklet containing phrases related to the sentences they had just heard. They were asked to write out the full sentence that each phrase reminded them of most. The experimenter then distributed the prompt booklets
and paced the subjects at 40 sec per prompt.

Results

Sentences were scored correct if a subject recalled both the topic and the vehicle. A topic or vehicle was considered correct if it included the central noun from the original topic or vehicle noun phrase. Paraphrases were accepted if close synonyms were substituted for topic or vehicle terms and if the order of topic and vehicle was reversed.

The mean proportion of sentences correctly recalled by subjects is recorded in Table 1 for each condition. Recall with topic and vehicle prompts was nearly perfect. This is not surprising since a topic or vehicle prompt supplies half of the sentence that must be recalled. However, it does indicate that nearly all of the sentences are available to subjects for later recall. Thus, these recall scores suggest an upper limit on how well recall might be prompted under the best conditions.

<table>
<thead>
<tr>
<th>Acquisition list</th>
<th>Prompts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Topics</td>
</tr>
<tr>
<td>A</td>
<td>.86</td>
</tr>
<tr>
<td>B</td>
<td>.86</td>
</tr>
</tbody>
</table>

The results for ground prompts showed a strong interaction between Lists (A and B) and Grounds (A and B); this was verified in an analysis of variance for those four conditions [Lists x Grounds, F(1,60) = 146.8, p < .001]. There was no main effect for either Lists [F(1,60) = 0.85] or Grounds [F(1,60) = 0.05], suggesting that the lists were evenly balanced with respect to ease of recall and effectiveness of their related grounds in prompting recall. The source of the interaction is clear in Table 1: the grounds were effective as prompts only when subjects had heard the relevant acquisition sentence. In the case of Grounds A, recall of List A was far superior to recall of List B [F(1,60) = 62.6, p < .001]. With Grounds B, recall of List B was far superior to recall of List A [F(1,60) = 85.0, p < .001]. Similarly, from the standpoint of each acquisition list, relevant ground prompts were far more effective than grounds that were true of the topic but irrelevant to the sentence heard [F(1,60) = 76.2, p < .001 for List A; F(1,60) = 70.6, p < .001, for List B]. Overall, relevant grounds enabled subjects to approach perfect recall; recall was not far below the levels found for topic and vehicle prompts.
In this analysis, the matched pairs of acquisition sentences provided an internal control on the effectiveness of the grounds as prompts. Analysis of the variability associated with subjects showed that subjects performed best when the set of grounds was relevant to the set of acquisition sentences. Because of the crossed design of lists and prompts, the ineffectiveness of a set of irrelevant grounds could not be due to some flaw intrinsic to the prompts themselves (whether accidentally or by design): the same prompts were very effective when acquisition conditions were favorable. It is important to know for how many acquisition sentences and how many prompts this was true. If the effect was contributed by only a fraction of prompts that worked exceptionally well in relevant list conditions, then the results for subjects would be of far less interest.

To make an analysis of the behavior of prompts, we derived new scores from the original data by summing the number of subjects correctly recalling a sentence in each condition. The initial head count was impressive: 26 of the 28 acquisition sentences were better recalled with the relevant prompt than with the irrelevant prompt, and all of the 28 grounds were more effective in prompting recall of the relevant acquisition sentence.

To make a stronger test of these differences, we performed an analysis of variance for the behavior of prompts analogous to that performed above for the behavior of subjects. (We chose to study the variance associated with prompts, rather than acquisition sentences, since prompts were likely to show more variability and could be considered a repeated measure across list conditions, in each case providing a more sensitive test.) The mean proportions of subjects correctly recalling a sentence are equivalent to those in Table 1. Scores for the topic and vehicle prompts showed low variance, which verifies our earlier conclusion that all of the sentences are available for recall under optimal conditions. In the ground prompt conditions, there was again no main effect for either Lists [$F(1, 26) = 0.67$] or Grounds [$F(1, 26) = 0.03$], but there was a strong interaction between them [$F(1, 26) = 115.7$, $p < .001$]. The source of the interaction was clear: prompts performed best when subjects had heard the relevant metaphors. All within-level cell mean contrasts in the Lists X Grounds matrix were significant. For each list, relevant prompts were superior [$F(1, 26) = 36.3$, $p < .001$, List A; $F(1, 26) = 33.6$, $p < .001$, List B]. For each prompt set, more subjects recalled sentences in the relevant list [$F(1, 26) = 9.02$, $p < .01$, Grounds A; $F(1, 26) = 12.2$, $p < .01$, Grounds B]. Thus, we can reject the hypothesis that the high recall in relevant prompting conditions was attributable to only a subset of prompts that (fortuitously or not) produced high recall. The results were general for each set of prompts. (We might add that the distribution of scores for each set showed no bimodality.)

A few prompts produced high recall of the related irrelevant sentence. For example, out of 16 subjects in the List A/Grounds B condition, 9 correctly recalled the skyscraper-honeycomb sentence when given the irrelevant ground, are very tall compared to surrounding things. This is apparently a case where the ground is so criterial a property of the topic that it is likely to remain invariant and salient no matter what the context of interpretation (see Lakoff, 1972). However, as the above analysis makes abundantly clear, such cases were exceptions to an otherwise consistent pattern: topics interpreted
in one context tended to be inaccessible from other contexts.

Discussion

The results demonstrate that an abstract statement of the implicit ground of a metaphor is sufficient to remind a person of the metaphor at some later time. These abstractly related grounds were nearly as effective in prompting recall as the topics and vehicles explicitly mentioned in the sentences. The results are consistent with the hypothesis that subjects infer a resemblance during their initial encounter with a metaphoric sentence and that the resemblance is integral to what is stored as a memory of that experience. The interaction between lists and grounds further suggests that the semantic role of the topic is highly specific to the context supplied by the vehicle.

Before accepting these conclusions, however, there are other interpretations that must be considered. These are some of the alternatives, including one to which this study was directly addressed.

(i) Topic-property recognition. The vehicle does not interact in any way with the topic. The relevant ground is a (more or less salient) property of the topic. It prompts recall because the same property was activated during acquisition and formed part of the record of the event, or because in scanning a record of topics plus vehicles during recall, the system notes a match between the ground and the predicates or features already associated with the topic. According to this view, the vehicle is carried along as baggage, like the second term in a paired associate. A ground, therefore, should be as effective in prompting recall of "irrelevant" sentences (with the same topic) as it is in prompting recall of the relevant sentence. The experiment just reported indicates that this extreme position is untenable: the particular vehicle to which a topic is paired makes a substantial difference. Experiment II explores a more sophisticated version of this model.

(ii) Vehicle-property recognition. The relevant ground is more likely to be a salient property of the vehicle than of the topic. It prompts recall because the same property was activated by the vehicle during acquisition and formed part of the record of the event, or because in scanning through a record of topics plus vehicles during recall, the system notes a match between the ground and the predicates or features already associated with the vehicle. In this case, properties of the vehicle are seen as central to recall. The topic is carried along as baggage, and our understanding of it need in no way be transformed or enhanced. The particular topic to which a vehicle is paired should make little difference in the effectiveness of the relevant ground as a prompt for recall. This possibility will be tested in Experiment III, along with the possibility that the topic's or the vehicle's properties may provide the path for recall.

(iii) Topic or vehicle generation-recognition. Independent of any experience with the sentences, the likelihood is high that the relevant ground will make people think of the topic or the vehicle. It is sufficient to hypothesize that listeners make a kind of paired-associate record of the topic-plus-vehicle inputs, that the grounds lead them to generate many possible topics and vehicles at a later time, that they search their input
list for a sentence that contained the generated term, and that they then output any sentence where they recognize a match. According to this view, one need not assume that properties of either the topic or the vehicle are activated at acquisition or compared at recall. It is an extreme form of the proposition that the topic and vehicle do not interact in any significant way. We will test this possibility in Experiment IV.

EXPERIMENT II

The results of the first experiment suggest that relevant acquisition experience facilitates the effectiveness of the grounds as prompts, and that irrelevant or conflicting experience interferes with their effectiveness. Therefore, one might propose a more dynamic version of the topic-property recognition model in which properties of the topic are primed or weighted differently in the presence of different vehicles. This involves relaxing the rather extreme constraint that the vehicle not interact with the topic, but preserves the assumption that pre-existing predicates or features of the topic are the basis for interpretation and recall. Prompted recall with relevant grounds would presumably be effective because those properties of the topics were specially primed, tagged, or weighted during acquisition. This change might coincide with a reduced weighting being given to other properties (including the irrelevant ground) and, in any case, it would presumably affect the recognition of other properties during recall.

This model can become very alluring, so we must keep its potential failings clearly in mind. The "property" under discussion may not be part of a person's knowledge before hearing a metaphor, and, even if it is familiar, it may have to be rediscovered with a nuance unique to that context. Metaphor not only brings us to see the unfamiliar, but to see the familiar in new ways. The process of comprehension may involve more than activating a relatively stable network in a novel way or priming an unusual subset of features. It may involve a restructuring of the topic domain. Such a novel structuring would allow one to apprehend certain relations with ease, while other possible relations would be unavailable because apprehending them presupposes a different structuring. For example, the two metaphors about tree trunks invite us to structure our conception of tree trunks in entirely different ways. In contrast to the straws metaphor, the pillars metaphor leads us to conceive of trees as solid columns (rather than hollow tubes), to conceive of a forest of trunks (rather than an individual trunk), and to conceive of their function as holding up a solid mass of leaves and branches (rather than as transporting liquid to more individuated leaves and branches). We are not dealing with the same tree trunks in the two sentences, even though the isolated lexical items are identical.

In this experiment we studied the effect of the metaphoric vehicle by comparing subjects' interpretations of a topic with and without a vehicle.

3W. J. J. Gordon (1961), in his application of metaphoric thinking to problem solving, describes these functions epigrammatically as "making the strange familiar" and "making the familiar strange."
Subjects' recall of a list of isolated topics was prompted with the two sets of grounds, providing a measure of "comprehension" when no vehicles had affected interpretation of the topics. The overall design crossed three acquisition lists (A, B, and Topics Only) with two ground sets (A and B). Thus, for each set of grounds the following predictions could be tested.

(1) A ground should be less effective in prompting an isolated topic than in prompting a full metaphor with the relevant vehicle. This prediction would follow from any model that proposes interaction of the vehicle with the topic.

(2) A ground may be more effective in prompting an isolated topic than in prompting a full metaphor with an irrelevant vehicle. This prediction would follow if there is a greater likelihood that subjects will hit upon the "correct" context or properties while thinking about the isolated topic, compared to the topic in a conflicting context.

Method

A third acquisition list, List Topics Only, was recorded according to the same procedures used in recording Lists A and B. The list contained the topic noun phrases from the metaphors in the full-sentence lists, in the same order of appearance. Each topic was spoken twice, followed by a 5-sec pause. Two sets of prompt booklets, Grounds A and B, were identical in design to those used in Experiment I.

Subjects were 60 undergraduates in an introductory psychology course at the University of Minnesota. They received extra credit for their participation. Subjects were randomly assigned to one of three acquisition conditions: List A, B, or Topics Only. Ten subjects in each condition received Grounds A, while ten received Grounds B.

The procedure was the same as before. The experimenter read the acquisition instructions, and the subjects then listened to one of the three lists. In the List Topics Only condition the instructions were modified slightly; subjects were told they would hear a "short series of words and phrases" and were asked to "think about what each word or phrase is describing." Recall instructions were the same as before, except that subjects were asked to write down just the "topic" or "subject" of the original sentence (Lists A and B), or the "word or phrase" from the original list (List Topics Only). Thus, the recall tasks of all three groups were equalized to the extent that all subjects were to recall a phrase of equal length, and all responses could be scored according to the same criteria. Following the recall instructions, the experimenter distributed the prompt booklets and paced recall at 25 sec per prompt (compared to 40 sec in Experiment I, where both topic and vehicle were to be recalled).

Results

Topics were scored correct according to the same criteria used for accepting topics in full-sentence recall, that is, the response had to contain the central noun from the original topic noun phrase or a close synonym.
The mean proportion of topics correctly recalled by subjects is recorded in Table 2 for each condition. The pattern of recall in the two full-sentence list conditions replicates the pattern found in Experiment I. The level of recall for each group is also essentially the same as in the earlier experiment. Thus, it makes little difference to subjects whether they are asked to recall just the topic or the topic plus vehicle. If they can recall the topic, they will also be able to recall the vehicle with which its interpretation was (we presume) intimately connected.

<table>
<thead>
<tr>
<th>Prompts</th>
<th>Grounds A</th>
<th>Grounds B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition list</td>
<td>A</td>
<td>.69</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>.21</td>
</tr>
<tr>
<td>Topics Only</td>
<td></td>
<td>.41</td>
</tr>
</tbody>
</table>

An analysis of variance for the six treatment groups in Table 2 showed no main effect of either Lists [F(2,54) = 1.45] or Grounds [F(1,54) = 0.50], but there was a large interaction between the two factors [F(2,54) = 51.7, p < .001]. One source of this interaction is familiar from Experiment I: relevant pairings of prompts with full-sentence lists produced high recall; irrelevant pairings produced low recall. For each acquisition list, relevant grounds were much more effective as prompts [F(1,54) = 47.3, p < .001, List A; F(1,54) = 56.1, p < .001, List B]. For each prompt set, relevant acquisition experience was far superior in facilitating topic recall [F(1,54) = 67.7, p < .001, Grounds A; F(1,54) = 37.7, p < .001, Grounds B].

A second source of the interaction is clear in a comparison of the recall for full-sentence lists and for the topics-only list. For each set of grounds, recall of the isolated topics was intermediate between recall of the same topics in the context of relevant vehicles and recall of the topics in the context of irrelevant vehicles. With Grounds A, recall for List A was superior to recall for List Topics Only [F(1,54) = 22.9, p < .001], which in turn was superior to recall for List B [F(1,54) = 11.8, p < .01]. With Grounds B, subjects more successfully recalled the topics of List B than the same topics in List Topics Only [F(1,54) = 11.8, p < .01], which in turn were better recalled than the same topics in List A [F(1,54) = 7.30, p < .01].

Again, the generality of these findings needs to be verified by an analysis of the behavior of individual prompts. We need to be sure that the results are not the fortuitous contribution of a small subset of the
metaphoric sentences and their grounds. An analysis of variance (taking prompts as a repeated measure across lists) showed no main effect for either Lists [F(2,52) = 0.94] or Grounds [F(1,26) = 0.12], but there was a large interaction between the two factors [F(2,52) = 33.3, p < .001].

The results for the full-sentence list conditions verified those found in the analysis of prompts in Experiment I. For each list, relevant grounds enabled more subjects to recall the appropriate topic [F(1,26) = 11.7, p < .01, List A; F(1,26) = 13.9, p < .001, List B]. For each set of grounds, relevant acquisition experience was superior in facilitating correct recall [F(1,26) = 11.9, p < .01, Grounds A; F(1,26) = 6.60, p < .025, Grounds B].

A second source of the overall interaction was found in contrasts between full-sentence and topics-only lists. However, the results of these contrasts were not as clear-cut as they were in the analysis by subjects. With Grounds A, the intermediacy of isolated topics between relevant and irrelevant metaphors was significant, but not strongly so. Topics of metaphors in List A were definitely recalled by more subjects than the same topics in List Topics Only [F(1,26) = 9.84, p < .01], but these in turn were only somewhat better recalled than the same topics in the metaphors of List B [F(1,26) = 4.30, p < .05]. With Grounds B, for which within-group variances were especially high, the intermediacy of isolated topics was even less sharply defined: List B was better recalled than List Topics Only [F(1,26) = 4.30, p < .05], but List Topics Only was only marginally better recalled than List A [F(1,26) = 3.13, .05 < p < .10].

Study of individual prompts verifies the inconsistency of their behavior and suggests the source of any observed intermediacy of topics-only lists between relevant and irrelevant lists. While 24 (or nearly all) of the 28 grounds produced better recall of topics from relevant metaphors than from irrelevant metaphors, the recall of isolated topics was intermediate between the two in only 13 of the total cases. (We considered intermediacy to be any case where the number of subjects recalling a topic met the following inequality across the three list conditions: relevant list > topics-only list > irrelevant list.) The scores for individual prompts in the topics-only list condition showed modal values of three or four subjects recalling the topic, but the scores were spread throughout the range from 0 to 10 subjects. Most of the extreme cases happened to be prompts in Grounds B, accounting for the high variance in that condition. High recall apparently occurred when the ground was a salient or criterial property of the topic. For example, 10 out of 10 subjects recalled the isolated topic skyscrapers in response to the ground are very tall compared to surrounding things. (Recall that this was also an effective prompt for the "irrelevant" skyscraper-honeycomb sentence.) Low recall of isolated topics occurred when a ground required a relatively novel context for interpreting the topic. In response to the ground are tubes which conduct water to where it's needed, no subjects recalled the isolated topic tree trunks. Similarly, no subjects recalled billboards in response to the ground tell you where to find businesses in the area. Apparently, the likelihood was very low that they would think of the relevant context during their original contemplation of the topic, or the likelihood was low that they could see the ground's relevance to the topic even if they scanned over the topic during recall. The power of a vehicle to lead subjects to discover this
relevance is apparent in the recall scores for these same grounds in relevant list conditions: 6 out of 10 subjects correctly reported the topic tree trunks (having earlier heard the tree trunks-straws sentence), and 8 out of 10 subjects correctly recalled billboards (having heard the billboards-yellow pages sentence). The intermediacy in the topic-only list conditions apparently represents a central tendency along a continuum of likelihood that the relevance of a property will be noted in a null context.

Discussion

The results in the full-sentence list conditions support the claim that the vehicle plays a critical role in the comprehension and recall of metaphoric topics. If all properties of a topic could be activated at acquisition or recall, then any of them should serve to remind subjects of the topic. This was clearly not the case. With few exceptions, a specific property was a successful prompt only if it was integral to comprehension of the full sentence. When it was not integral to comprehending the sentence, subjects were only occasionally able to see its relevance to the topic at a later time.

The results in the topics-only list condition support this conclusion. If all possible properties were activated to an equal degree whenever the topic appeared, there should be no difference between isolated topics and any full metaphor containing them. But there were consistent differences; a particular property tended to be a good prompt for a relevant metaphor, variably intermediate for the topic alone, and a poor prompt for an irrelevant metaphor. Moreover, there was little correlation between the perceived relevance of a property to an isolated topic and its perceived relevance to the topic in context, as measured by prompted recall in each case. Across the 28 grounds, the correlation between the number of subjects recalling the topic from the relevant-sentence list to the number recalling it from the topics-only list was only 0.23.

These results do not support a simple form of the topic-property recognition model. A more sophisticated form of the model would need to propose how the vehicle enhances the saliency of one or more of the topic's properties. Models written in the framework of semantic feature theory and semantic network theory typically propose a search for common features or common associations (including associated predicates). For example, Johnson et al. (see footnote 2) and Malgady and Johnson (1976) compare metaphors to compound association stimuli and argue that features shared by the two nouns are raised in saliency, compared to non-overlapping features. They report that rated "figure goodness" correlates with the degree of rated similarity between the two nouns and the number of (independently assessed) shared attributes. Sternberg (1977) proposes that judgments about the validity of four-term analogies are based on component processes that include scanning for feature matches. Similar accounts can be written in terms of overlapping activation of predicates in a semantic memory model. Kintsch (1972, 1974), for example, suggests that the meaning of a metaphor is based on common "lexical implications" associated with its underlying terms.
All of these approaches assume that the ground of a metaphor is the logical intersection of two pre-existing sets of semantic elements, and that a sufficient comprehension strategy is to search for these common elements. An all-too-easy inference from these models is that sentences linking highly similar things in familiar contexts are quintessential metaphors: Skyscrapers are the giraffes of a city, and even Flowers are the blooms of a garden. Clearly, such a similarity continuum provides no basis for distinguishing metaphoric language from literal language or tautology, let alone for characterizing aesthetic quality.

While the common-elements approach appears to handle the most transparent comparisons, it is inappropriate for most of the sentences in this study. Properties that were poor prompts of the isolated topics cannot reasonably be said to be low-frequency or low-saliency entries in a pre-existing set of the topic's properties. We only become aware of such properties when a particular vehicle invites us to do so. We can add these properties, post hoc, to our list, but we will never be able to specify exhaustively all of the resemblances that we may potentially discover. Many studies of metaphor and analogy beg this question by using small preselected sets of attributes and values, and by making their identity obvious to subjects from the outset (for example, Sternberg, 1977). In natural contexts of metaphor or analogy use, the crucial task of comprehension is to discover what properties are relevant. The vehicle certainly plays a role in determining what is "relevant," but these constraints cannot be modeled effectively by a weighted matching function that selects out pre-existing attributes of the topic. As an account for all of the metaphors studied here, it may prove more parsimonious to say that "priming" results from a distinctive structuring of the topic domain for each metaphoric context in which the topic terms appear.

**EXPERIMENT III**

To this point we have considered properties of the topic as the focal point for processes in recall. The simple topic-property recognition model received negligible support. The specific vehicle paired with a topic exerts considerable influence on the topic's interpretation and its accessibility to recall at a later time. In cases where the ground is not part of prior knowledge about the topic, the vehicle's role in defining sentence meaning is clearly central. This leads us to consider a second possible class of featural explanations for the high level of prompted recall in relevant list conditions: vehicle-property recognition. In many cases the relevant ground is a salient property of the vehicle (considered in isolation). The use of such a vehicle presumably makes the metaphor more comprehensible and more effective in attributing a property to the topic. For example, the ugly protrusiveness of warts and the tallness of giraffes are both salient properties. The relevant grounds may be effective prompts because they specify properties that are activated when hearing the vehicle at acquisition, or that are easily discovered during some scanning process at recall.

There are various forms this hypothesis could take. Linguists and rhetoricians have often asserted that metaphor involves a transfer of meaning from the vehicle to the topic. (The Greek ancestor of the term "metaphor" meant to transfer or carry over.) In recent attempts to accommodate feature
theory to metaphoric language, semantic interpretation is described as a
transfer of part of the feature specification of the vehicle to the topic,
adding and altering values in the feature specification of the topic
(Weinreich, 1966; Bickerton, 1969; Leech, 1969; Thomas, 1969). In linguistic
terms, this usually constitutes a more-or-less temporary alteration in the
dictionary entry for the topic. A similar process could be proposed in the
framework of semantic memory models: the transfer would consist of adding a
new predicate to the current representation of the topic. Orthodox
behaviorists and mediationists might argue that metaphor is simply a case of
classical conditioning. By pairing the topic and vehicle in close temporal
contiguity, the ground (which is a strong unconditioned meaning response to
the vehicle stimulus) may be transferred to the topic stimulus (see Osgood,
1953; Mowrer, 1954).

For each of the strong forms this hypothesis can take, the same
conclusion follows directly: prompting of recall should be equally effective
no matter what topic a vehicle is paired to, since the vehicle's properties
determine the meaning and are the focal point for processes in recall. For
the sentences in Experiments I and II, the vehicles were chosen to make
comprehensible assertions about the topics (we will call these "principled
metaphors"). The vehicle-property hypothesis suggests that the specific
pairings should make little difference. Therefore, for this experiment we
randomized the pairings of topic and vehicle phrases to create a new set of
metaphoric sentences ("arbitrary metaphors"). If the relation of the vehicle
to a ground is all that determines recall, then recall of these new metaphors
should be as high as recall for the original metaphors. Only "relevant" list-
gounds pairings were used in this experiment, for comparison with relevant
prompted recall conditions in Experiment I.

Method

Two acquisition lists of arbitrary metaphors (Lists A' and B') were
prepared from the principled metaphors by randomly reassigning pairs of
vehicles to different topics. For example:

Tree trunks are like dragons.
Tree trunks are like babies with pacifiers.

Cigarette fiends are warts on the landscape.
Cigarette fiends are the yellow pages of a highway.

Note, however, that metaphoric interpretations vary widely in permanency.
Some metaphors request only a short-term orientation to a topic, as in the
comparison of tree trunks to straws. Others presuppose more permanent (and
more global) modes of orienting to the environment; for example, a tree trunk
may be viewed as the residence of a malevolent being or as the umbilical of
the Great Earthmother in a myth of biological genesis (Keeler, 1961). The
duration of a metaphoric interpretation is another aspect of metaphor use
that cannot be accounted for in terms of a user-independent axiomatic
semantics.
The order of topics in each list was the same as in the comparable lists of principled metaphors (Lists A and B). The singularity/plurality of the topic and verb was adjusted in some cases to correspond to that of the vehicle. With this minor exception, the new lists contained the same verbal material as the original lists; thus, the memory tasks (simply conceived) and the possible intralist confusions were comparable. The lists were recorded under the same conditions as before; the intonation contours and pace were kept as natural as possible. Each sentence was repeated twice and was followed by a 5-sec pause.

The prompt booklets were identical in design to those used before (Grounds A and B). Thus, the order of correct recall of vehicles (and the topics paired to them) was the same.

Subjects were 20 undergraduates enrolled in an introductory psychology course at the University of Minnesota. They received extra credit for their participation. Subjects were randomly assigned to one of two conditions: 10 subjects heard List A' and received Grounds A as prompts, and the other 10 heard List B' and received Grounds B.

The listening conditions and acquisition instructions were the same as before. The experimenter mentioned that some of the sentences would be a little bizarre and asked subjects to do their best to find sensible interpretations. Recall instructions were those used in Experiment 1, that is, subjects were asked to recall the full sentence most related to each prompt, as well as they could remember it. They were paced at 40 sec per prompt.

Results

In scoring subjects' responses for the appearance of topics and vehicles, the same criteria were used as in previous experiments. In the initial scoring procedure, the sentence containing the vehicle originally related to the ground was judged to be the "correct" sentence to recall. Both the topic and the vehicle of this sentence had to be correctly recalled.

The mean proportion of arbitrary metaphors recalled per subject is recorded in the second column of Table 3 for each list condition; the results for principled metaphors in comparable conditions are included in the first column for comparison. The results were clear: when a vehicle appeared in a principled metaphor, relevant prompted recall of the sentence was substantially greater than when the same vehicle appeared in an arbitrary metaphor. This difference was significant for both sets of grounds [two-tailed $t(24) = 4.04$, $p < .001$, Grounds A; $t(24) = 5.53$, $p < .001$, Grounds B]. This rules out any simple hypothesis that ascribes relevant prompted recall solely to the relation between the ground and the vehicle.
TABLE 3: Mean proportion of sentences recalled: Experiment III.

<table>
<thead>
<tr>
<th>Prompts</th>
<th>Principled metaphors</th>
<th>Vehicle sentence</th>
<th>Topic sentence</th>
<th>Topic or Vehicle sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grounds A</td>
<td>.70</td>
<td>.40</td>
<td>.11</td>
<td>.51</td>
</tr>
<tr>
<td>Grounds B</td>
<td>.73</td>
<td>.34</td>
<td>.16</td>
<td>.50</td>
</tr>
</tbody>
</table>

\(^{a}\)From Table 1, Experiment I.

In the previous experiments we considered a complementary hypothesis that ascribed recall solely to the relation between the ground and the topic. The present results allow another test of that hypothesis. Subject's responses were rescored, counting as "correct" any sentence that contained the topic originally related to each ground prompt. The mean proportion of sentences correctly recalled by subjects according to this criterion is recorded in the third column of Table 3 for each condition. A sizable fraction of the arbitrary metaphors correctly recalled by subjects resulted from a close relationship between topics and grounds. Even so, the fraction attributable to topics was substantially smaller than that attributable to vehicles. Topic-property recognition is even less successful than vehicle-property recognition as a predictor of the level of recall for principled metaphors.

We are now in a position to test a combined hypothesis: the recall of metaphors may involve prompting of either the topic or the vehicle (by means of an associated property that matches the ground), followed by recall of the other member of the pair. A comprehension process laying the groundwork for this recall process could be framed in terms of probabilities or saliencies. There may be a certain probability that an appearance of the topic will activate a relevant property, and an independent probability that the vehicle will activate the same property. There may be a certain resting saliency of the property in the topic domain and an independent saliency in the vehicle domain. The possible success of a combined hypothesis is suggested by results for some of the arbitrary metaphors. In the few cases where a topic sentence was frequently recalled, the ground tended to be a salient property of the topic; for example, 4 out of 10 subjects recalled the skyscraper-branding iron sentence in response to are very tall compared to surrounding things. In cases where a vehicle sentence was frequently recalled, the ground tended to be a salient property of the vehicle; 9 out of 10 subjects recalled the cigarette fiends-warts sentence in response to are ugly protrusions on a surface.

Whether the combined model is phrased in terms of prior probabilities or saliencies, the critical assumption is that the values associated with the topic and vehicle domains are independent. If probabilities related to the
vehicle are zero, the model reduces to a topic-recognition model. If probabilities related to the topic are zero, we have a vehicle-recognition model, and it is irrelevant whether we choose to speak of "transfer" of properties to the topic. If both probabilities are nonzero, we have the model described at the end of Experiment II: the ground of a metaphor is the intersection of two independent property sets. The relation between the ground and the metaphor will be characterized by a joint probability in addition to the probabilities associated with the topic and vehicle alone. This model, in the language of saliencies, is best exemplified by the work of Johnson et al. (footnote 2) and Malgady and Johnson (1976).

The combined model asserts that the probability of recall of principled metaphors is the sum of the probabilities for prompting only topic recall, only vehicle recall, and both topic and vehicle recall. (This assumes that the probability is unity of getting from only the topic or the vehicle to the full sentence; the results of Experiment I indicate this is a reasonable assumption.) The recall data for arbitrary metaphors do not allow us to estimate these three probabilities directly, since we do not know how subjects divided their responses between the topic and vehicle sentences when both came to mind. However, we can estimate the total probability by summing the topic and vehicle sentences recalled by each subject and averaging the new set of scores. These estimates are recorded in the fourth column of Table 3. For each set of grounds, the mean for topic or vehicle sentence recall was significantly less than the mean for principled sentence recall [$t(24) = 2.33$, $p < .05$, Grounds A; $t(24) = 2.80$, $p < .01$, Grounds B]. In addition, at the level of individual prompts, there was no correlation between the frequencies of recall for arbitrary and principled metaphors ($r = 0.005$ for the 28 grounds). Thus, a combined model, assuming independently defined probabilities or saliencies for the topic and vehicle, is not adequate as a predictor for the recall of metaphoric sentences and, by implication, may not be adequate as an explanation for their comprehension.

Discussion

It is possible to accept this conclusion without negating the intuitions that motivated the models tested here. For example, the importance of salient aspects of the vehicle domain seems unquestionable. The vehicle exerts a tremendous influence on the accessibility of principled metaphors to recall, and it is clearly the more common pathway for recall of arbitrary topic-vehicle combinations. Thus, the comprehension of metaphor may involve a presupposition that the dominant source of constraints on meaning is the vehicle, and that the topic should be comparatively malleable to interpretation. Even if one argues for a mutual influence of topic and vehicle domains on each other, it seems clear that the degree of influence is asymmetrical.

5It should be noted that almost all of the sentences correctly recalled were either topic sentences or vehicle sentences. Thus, the lower total recall for the arbitrary metaphors cannot be attributed to the intrusion of incorrect responses. The number of intrusion errors in Experiment I was similarly small.
This again raises the question of independence and interaction. With the exception of the more extreme vehicle-property transfer theorists, almost everyone would agree that the topic and vehicle "interact" in a comprehender's interpretation of metaphor, in the loose sense that both affect the resulting meaning. There are two levels, however, at which the question of independence needs to be posed. At the more fundamental level, we must ask whether the topic and vehicle are "separable." This is a question about what hypothetical entities provide the most useful basis for an explanatory theory of the process of comprehending metaphor. If we assume the topic and vehicle to be separable, then we are assuming that they have associated properties, probabilities, saliencies, states, or processes that are independently defined. Having assumed distinct entities at this level, we can proceed to ask whether the two sets of entities interact in the hypothetical processes underlying comprehension. Most of the current linguistic and psychological approaches to semantic interpretation assume separability of the entities attributed to individual words: their features, concepts, predicates, meanings, associations. For example, Johnson et al. (see footnote 2) attribute distinct feature vectors to each term and then define the meaning of the full metaphor in terms of the union and intersection of these two feature vectors. They make a point of asserting that this is an "interactive" process, and, in a secondary sense, it is; but at the fundamental level their model assumes that the two terms function independently and additively. A comparable distinction would apply to semantic network accounts of metaphor; these models assume separate storage of information for each domain and define semantic interpretation in terms of new interconnections.

The assumption of separability is a natural one. We perceive words and objects as having separate identities, and it is natural to try to characterize these identities in isolation. Dictionaries serve useful functions, and it is tempting to assume that hypothetical dictionaries (lexicons or networks) will provide a sufficient base for hypothetical processes of comprehension. The crucial question for cognitive theory is whether words are functionally separable. In the pursuit of meaning, in response to sentences and longer discourse, the cognitive impacts of component words may be only partially separable.

The results for arbitrary metaphors provide a strong (though certainly not definitive) test of models assuming separability of words and a more-or-less additive process for their combination. To these models, all topic-vehicle combinations are fundamentally arbitrary. However, it is clear from the data that "arbitrary" pairings do not have the cognitive force of "principled" pairings (intuitively defined). Subjects' performance on arbitrary pairings did not provide adequate estimators for their performance on principled pairings. It is also worth noting that the frequency of recalling only a topic or a vehicle was substantially higher for arbitrary metaphors than for principled metaphors. Recalling the topic or vehicle of an arbitrary metaphor does not always allow recall of the other member of the pair; thus, the assumption made above that this probability is unity does not hold for arbitrary pairings. This suggests that subjects' representations of arbitrary pairings are less integrated; they have been forced to deal with many of the topics and vehicles as separate entities. One further symptom of this is the
appearance of combinations in recall that were not heard during acquisition. In response to are very tall compared to surrounding things, one subject responded with a sentence combining two topics: Skyscrapers are billboards to a large city. Another subject recombined two pieces to produce the original principled metaphor: Skyscrapers are the giraffes of a city. In addition, four subjects recalled the related topic sentence (skyscrapers-branding irons), one recalled the vehicle sentence (matches in a forest-giraffe), one recalled only the topic (skyscrapers), and two recalled only the vehicle (giraffes).

To a language user, the "same term" is not the same term in each context of combination. The "same vehicle" need not have the same predicating potential in all contexts. A predicate that is an effective prompt in one topic context (principled metaphors) need not be effective in another topic context (arbitrary metaphors). Similarly, the "same topic" is not functionally the same when combined with different vehicles. The possible relevance of a predicate to a topic may be perceptible only if the topic has appeared in the context of a particular type of vehicle. As argued above, this kind of flexibility in a term's function is true of all language use and cannot be characterized by prescriptions in a lexicon. The crucial question for metaphor is not what constraints need to be relaxed, but what constraints need to be imposed to make metaphoric combinations interpretable. The topic and vehicle are not totally flexible; arbitrary combinations are not as easily integrated as principled combinations. The reason for this may be the receptiveness of the topic to the "structuring" suggested by the vehicle (assuming the vehicle plays the dominant role). We can easily transform a tree trunk into a straw or a pillar, but not so easily into a dragon or a baby with a pacifier. It is doubtful that a logic of topic-vehicle compatibilities can be successfully framed in terms of elemental semantic features or predicates. The process of comprehension involves a more global transformation of the topic domain. Compatibility with a vehicle depends on the susceptibility of the entire domain to the appropriate transformation, and each such transformation defines new "properties" for the topic. It is in this sense that the topic's semantic structure is not fundamentally separable from the vehicle.

These considerations lead us to suggest that the comprehension process results in a partial identification (or fusion) of the topic and vehicle domains. To some extent, the imagined tree trunk may become a straw and the skyscraper may become a giraffe extending its neck above the city skyline. This mode of comprehension may be more common and integral to adult language use than is currently recognized. It has typically been assumed that "identification" is uniquely characteristic of pathological, poetic, or primitive thought; for example, the "paleologic" thinking of schizophrenics (as defined by Arieti, 1974), "primary process" thinking (for example, Freud, 1950), poetic imagination (Richards, 1960; Hawkes, 1972), symbolic play in children (Piaget, 1962; Gombrich, 1968), and magical thinking. While healthy use of metaphor does not typically entail a total identification of the topic and vehicle, the assumption of full functional separation seems equally extreme. Productive use of metaphor in problem solving, scientific theory, poetry, and personal growth probably demands a partial fusion of the two
domains.

EXPERIMENT IV

The models discussed in the previous experiments assume that particular properties are apprehended during the process of comprehension, and that they later determine the accessibility of the topics and vehicles. We now consider an alternative approach that resists postulating such properties as mediators and attributes recall to a "direct" relationship between the grounds and the relevant topics and vehicles. For example, the phrase are ugly protrusions on a surface might lead subjects to think of warts independent of any special acquisition experience involving inference, matching, pairing, or other postulated processes. Prompted recall could consist of generating possible terms (for example, warts) in response to the prompt, searching some record of the original sentences until a matching term is recognized, and then reporting the sentence containing it. This recall procedure is similar to the "generation-recognition" model tested by Tulving and Thomson (1973) in their analysis of prompted recall for word lists, and it has been suggested by Osgood as a possible explanation for the data reported here. In its simplest form, the model treats a metaphor as an uninterpreted paired associate that is stored in an "episodic memory" (Tulving, 1972) for later recall. While this is not a satisfying explanation of what it means to understand a metaphor, it could be sufficient to account for our earlier data in relevant prompted recall conditions.

To test this possibility we need an estimate of how likely people are to think of the relevant topic or vehicle when they read a ground without any prior experience with the acquisition sentences. To make these estimates we devised the following sentence completion task.

Method

Two sets of mimeographed response booklets, Grounds A and B, were prepared. They contained the grounds for Lists A and B, respectively. A cover sheet informed subjects that their booklets contained some incomplete sentences. They were asked to complete each sentence by supplying a "subject," using either a single word or an extended phrase. They were asked to write down at least three possible subjects and to work quickly, recording their answers as soon as they came to mind. The following example was provided.

_________ are very colorful.

1. Flowers
2. Hawaiian shirts
3. Eccentric people

_________

The order of the phrases in each form was the same as in the prompt booklets used in earlier experiments.

Subjects were 64 undergraduates enrolled in introductory psychology courses at the University of Minnesota. They were randomly assigned to one of two groups, receiving Grounds A or B. Approximately half of each group received extra credit for their participation; the remainder completed the form as a class assignment. Subjects worked individually in a quiet experimental room or classroom.

Results

Responses to each ground were scored as "topics" or "vehicles" if the terms were identical to or closely synonymous with terms in the original topic and vehicle phrases of the relevant metaphor. For example, moles and pimples were also accepted for the vehicle warts; beehives was accepted for honeycombs; and IDS building (the skyscraper in Minneapolis) was accepted for skyscrapers. Separate tallies were made for topic and vehicle responses; only the first appropriate response of each type was recorded.

The mean proportion of topics and vehicles produced by subjects is recorded in Table 4 for each set of grounds. On the average, subjects were more likely to think of related vehicles than topics by a factor of about 2:1. This bias toward vehicle responses is similar to that observed in Experiment III and suggests a complementary hypothesis about why particular vehicles are chosen as metaphoric predicates: they are exemplary instances of particular relationships. When encountering a ground under free association conditions, subjects are more likely to think of the vehicle domain (where the relationship is familiar) than the topic domain (where its relevance may not be familiar).

<table>
<thead>
<tr>
<th>Set of grounds</th>
<th>Topics</th>
<th>Vehicles</th>
<th>Topics or vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grounds A</td>
<td>.05</td>
<td>.18</td>
<td>.22</td>
</tr>
<tr>
<td>Grounds B</td>
<td>.12</td>
<td>.17</td>
<td>.28</td>
</tr>
</tbody>
</table>

However, these domains are only two among many that are likely to come to mind. The question is whether they do so often enough to account for the level of relevant prompted recall in earlier studies. The third column in Table 4 records the mean proportion of topics or vehicles supplied by subjects.
for each set of grounds. On the average, subjects thought of 25 percent of the topics or vehicles. If being reminded directly of the topic or vehicle were a prerequisite for recall of principled metaphors, then we could expect subjects to recall no more than 25 percent of the 14 sentences, even if we assume recall proceeds without error once a topic or vehicle is known. This estimate falls far short of the level of relevant prompted recall observed in Experiment I, where subjects were able to recall about 72 percent of the sentences \( t(94) = 14.4, p < .001 \), for the two sets of grounds combined.

Not surprisingly, this finding is repeated in an analysis of grounds. For each ground in the sentence completion task, one can score how many subjects (out of 32) responded with the related topic or vehicle. The mean proportions of subjects are equivalent to the means in Table 4 and lead to a complementary conclusion: the probability that a topic or vehicle will be produced in response to a ground is substantially higher when subjects have heard the relevant acquisition sentence. This suggests a more sophisticated form of the generation-recognition hypothesis. The acquisition sentence may prime the topic and vehicle, making it more likely that they will be evoked during recall as implicit responses to the ground. If this priming is exerted equally by all topics and vehicles in the acquisition list, then the sentence completion data should enable one to predict the relative probability of prompted recall for individual grounds. For example, grounds that frequently evoke topic or vehicle responses in the sentence completion task should also produce high levels of correct recall in the prompted recall task. In other words, there should be a strong correlation between a ground's behavior in the two tasks.

A test of this hypothesis is facilitated by the substantial variability among grounds in each task. Experiment I measured the probability that each ground would produce correct recall of the full relevant sentence. We may take these as observed probabilities and test the power of an associative model to predict their configuration. Rough estimates of associative probabilities may be obtained from the proportion of subjects producing the topic or vehicle in response to each ground. These estimates assume that recall proceeds errorlessly if either the topic or the vehicle is implicitly generated.

Observed and estimated probabilities showed little systematic relationship. For the 28 grounds, the coefficient of correlation between these

---

7Inclusive or. Note that each figure is smaller than the sum of probabilities for topic and vehicle responses, since subjects occasionally responded with both. It is worth noting that the probabilities of responding with the topic and the vehicle are independent. The estimated probability of topic/vehicle co-occurrence would be \( (0.054)(0.176) = 0.0095 \) for Grounds A and \( (0.123)(0.174) = 0.021 \) for Grounds B. The mean observed probabilities of co-occurrence were not significantly greater than these estimates; the observed values were \( 0.0089 \) for Grounds A \( t(31) = 0.11 \), and \( 0.016 \) for Grounds B \( t(31) = 0.93 \). This suggests there was little or not pre-existing "associative strength" between the topics and vehicles of the original metaphors.
estimated probabilities and the observed probabilities was only 0.17. This comparison assumes that priming is a linear function of extra-experimental associative probability. If priming is assumed to preserve linearity of the logarithm of probability measures, the correlation remains low and nonsignificant \( r = 0.27 \). Thus, the associative model outlined above cannot successfully predict either the overall level or the specific configuration of relevant prompted recall.

More sophisticated probability estimates would acknowledge that recall may not proceed errorlessly if only the topic or the vehicle is generated. In Experiment I, there was some variability in the effectiveness of topic and vehicle prompts, and topics were slightly less effective overall than vehicles. A more accurate predicted probability for each ground could be obtained using the following equation:

\[
\hat{p} = p(T) \cdot p(S/T) + p(V) \cdot p(S/V) + p(TV) \cdot p(S/TV),
\]

where \( p(T) \) is the probability of responding associatively with only the topic, \( p(S/T) \) is the probability of producing the full sentence given the topic, \( p(V) \) is the probability of responding associatively with only the vehicle, \( p(S/V) \) is the probability of producing the full sentence given the vehicle, \( p(TV) \) is the probability of responding associatively with both the topic and the vehicle, and \( p(S/TV) \) is the probability of producing the full sentence given both the topic and the vehicle. Estimates of \( p(T) \), \( p(V) \), and \( p(TV) \) for each ground were obtained in this experiment (using a measurement scale of 32 subjects). \( p(S/T) \) and \( p(S/V) \) for each ground were obtained in Experiment I (using a much coarser scale of eight subjects). \( p(S/TV) \) may be assumed to be 1.00. Across the 28 grounds, the correlation of \( \hat{p} \) with the observed probability of relevant prompted recall was only 0.18. Thus, the more careful estimation procedure does not alter the original conclusion: the generation-recognition model cannot predict the configuration of prompted recall.

It is worth noting that in a few cases the original vehicle was a frequent response to the ground in the sentence completion task; for example, warts, pimples, and the like were common responses to are ugly protrusions on a surface \( (\hat{p} = 0.68) \), and yellow pages was a common response to tell you where to find businesses in the area \( (\hat{p} = 0.50) \). In one case the original topic was a common response to the ground: skyscrapers and IDS building were frequent responses to are very tall compared to surrounding things \( (\hat{p} = 0.69) \). In these exceptional cases, the original vehicles or topics happened to be the most salient instances of the relationship specified abstractly by the ground,

---

8Note that these estimates require rescoring the original data. Earlier we scored the number of subjects producing a topic or a vehicle (irrelevant of whether the other term co-occurred in individual subjects' responses). \( p(T) \) requires scoring responses which include only the topic, \( p(V) \) involves responses which include only the vehicle, and \( p(TV) \) is the probability of co-occurrence. This breaks down the earlier "rough" probability estimate (total topics or vehicles) into three components.
and the estimated recall probabilities approached the observed values. In general, however, responses to the grounds showed little correspondence (in either absolute or relative frequency) to the topics and vehicles produced in relevant prompted recall.

**Discussion**

The results of this experiment demonstrate that the hypothesis of pre-existing associations between grounds and topics/vehicles provides little explanatory power. Neither the overall level nor the specific configuration of recall can be accurately estimated from the strengths of such associations. At the very least, this confirms our intuition that recall of a metaphoric sentence cannot be ascribed to a direct prompting of component terms, but involves some kind of match between relationships experienced at the invitation of those terms and the relationships specified by the ground. The product of comprehension must be more than a novel paired associate, more than a new "link" between the two terms or two classes of objects.

Tulving and Thomson's (1973) discussion of paired-associate stimuli applies to some extent to the conjunctions of noun phrases in metaphors: while the "nominal memory unit" is no more than a conjunction of terms, the "functional memory unit" can be a much more elaborated cognitive product. It is the functional unit that governs accessibility of the terms to later recall. In the case of metaphor, the functional unit can be an elaborated event or structure in which the terms' referents are only local components. The relationship between the ground and this elaborated structure exerts a greater influence on recall than any pre-existing relationship between the ground and the particular components mentioned in the sentence.

The logic of this experiment was complementary to that of previous experiments, but led to similar conclusions. Models tested in the earlier experiments assumed the prior existence of stored predicates or features that would be activated during comprehension. These properties were assumed to provide a sufficient set of constructs for characterizing the resulting meaning and the possible entry points for recall. With few exceptions, the distinctive relationships between metaphors and grounds could not be explained satisfactorily by these models. In contrast to these models, which assumed strong "forward associations" between sentence terms and properties, the generation-recognition model tested in this experiment assumed strong "backward associations." Again, the distinctive relationships between metaphors and grounds could not be accurately predicted. The relationship created by metaphor has nothing necessarily to do with familiar ways of structuring knowledge.

To the extent that the strengths of the postulated forward and backward associations show some correspondence, this experiment could be viewed as a replication of Experiment III. The convergence of the two experiments is suggested by the similar distributions of topic and vehicle responses (see Tables 3 and 4) and the similar interaction with sets of grounds (A and B) in each case. It is possible that arbitrary metaphors more closely fit the assumptions of the generation-recognition model than principled metaphors.
There were suggestions that the terms in the arbitrary metaphors often did not interact in the specification of meaning, that the terms were more available for recall as isolated and interchangeable units, and that they were more likely to be interpreted in terms of normative properties. However, the failure to find a correlation between recall of principled and arbitrary combinations could have been due simply to the fact that different metaphoric combinations specify different grounds. The interpretation of arbitrary metaphors could be as novel and interactive as that of principled metaphors. If so, the sentence completion data should be no better as predictors for the arbitrary metaphors than they were for the principled metaphors. On the other hand, if the behavior of arbitrary metaphors is much more a consequence of normative properties of their component terms, then the estimated probabilities based on the sentence completion data may have greater predictive power.

Results suggest that prior associative connections play a much greater role in the recall of arbitrary metaphors. Across the 28 grounds, there was a significant correlation between frequency of topic responses (Experiment IV) and frequency of recall of topic sentences (Experiment III), \( r = 0.42, p < .05 \). The correlation between frequency of vehicle responses and frequency of recall of vehicle sentences was even stronger, \( r = 0.55, p < .01 \). Finally, we can consider the combined recall for arbitrary topic and vehicle sentences. The observed frequency of recall and the total estimated probability (\( \hat{p} \)) of recall showed a significant correlation, \( r = 0.48, p < .02 \). Thus, the results for arbitrary metaphors and free association to grounds are significantly correlated with each other, but neither set of results is closely related to the behavior of principled metaphors. Prior associative connections (whether forward or backward or both) apparently play little role in the comprehension and recall of nonarbitrary metaphoric sentences.

**GENERAL DISCUSSION**

These experiments gave no indication that metaphoric comprehension is a specialized skill in which only certain people excel, or that metaphoric sentences are especially difficult to comprehend. Our listeners showed no bimodality in recall performance, and their average level of recall in relevant prompting conditions was very high. If metaphoric comprehension is a skill in deviance, it is a normal one.

We have taken the high level of relevant prompted recall as evidence that listeners discerned an abstract resemblance between the topic and vehicle domains. A paraphrase of the ground was highly effective as a prompt, even though the resemblance was not explicit in the original sentence, and the prompt contained no content words from the sentence. The results of Experiment IV indicated the necessity of postulating this implicit resemblance as a central component of comprehension and a mediator for recall; direct associative connections between the prompts and acquisition sentences could not predict the configuration of prompted recall performance. Subjects' paraphrases in recall provided further evidence for the presence of these grounds in their interpretations. They occasionally added to or modified the original terms, making it clear that they had inferred the appropriate resemblance: Tree trunks are like straws that give drink to the leaves; Smokers are like
fire-breathing dragons.

These results have raised several issues concerning the structure of metaphoric resemblances, the process of comprehension, and the process of recall. In each case, we would like to sketch an alternative to attributive models that seems more consonant with our empirical findings and more fruitful as a vehicle for future theory and research. We hope this bold sketch will open avenues of investigation by which all models may become better articulated.

The Structure of Resemblance

In our discussion of the individual experiments, we considered various means of characterizing the grounds of metaphoric sentences. For both empirical and theoretical reasons, we have chosen to characterize metaphoric grounds in terms of abstract relations, rather than attributive features. We found negligible support for recall models that postulated the recognition of pre-existing attributes associated with topics, the priming or weighting of such attributes during acquisition, or the transfer of salient attributes associated with vehicles. While other models of this class could certainly be designed, we found no reason to believe that these were steps in the right direction.

A central question in this discussion is how the ground is related to the nominal terms of a metaphoric sentence. (We will limit ourselves here to sentences of the form "A" is (like) B," where A and B are both noun phrases.) Attributive models characterize the nominal terms by a list or array of features, and they characterize the ground by some weighted function of these features. These models are not well suited for characterizing grounds when the resemblance is not between the two terms (objects) per se, but between events or relationships in which each participates. Therefore, we prefer to describe metaphoric resemblances as relations between topic and vehicle domains (or schemata). Each domain is an abstract relationship among several entities; only a subset of these entities appears explicitly as nominal terms in the sentence. Thus, it is not strictly appropriate to identify the topic or vehicle of a metaphor with specific terms appearing in the sentence. In the tree trunks-straws sentence, for example, the topic term is tree trunks, but the topic domain is a type of transformation (fluid transport) exerted over certain structures (tree trunk, leaves and branches, water, roots, earth, etc.). A comparable description is also necessary for the vehicle domain, which is only partially specified by the terms straw and thirsty. The ground combines the transformational invariants (for example, suction, fluid flow) and structural invariants (for example, vertical cylindrical space) that are common to each domain.

A semantic characterization of nominal terms must be made in a way that facilitates achieving a topic domain, vehicle domain, and transformational/structural resemblances as the "product" of comprehension. Simply activating a set of normative, context-free, structural descriptors is not enough (inanimate, cylindrical, plastic, hollow, 6-10 in. long, etc.). It seems preferable to suppose that a nominal term can activate a system of
abstract structural and transformational invariants (that is, a domain or schema). These invariants will conjointly specify constraints on the relationships that the nominal term can participate in. The semantic characterization may also include particular instantiations of these abstract constraints within normative contexts. For example, the term straw could activate the following system of abstract constraints: a structure of relatively rigid nonporous material, of a hollow cylindrical shape, with a small diameter relative to its length. This structural specification is compatible with the accompanying transformational specification of event(s) within which the structure participates: the vertical cylindrical space channels fluid flow from a receptacle to a destination against gravity; the goal of the fluid transport is to alleviate thirst; the force for the flow is suction. In its normative contextual instantiation, the structure is paper or plastic, the receptacle is a bottle or cup, the destination is a person (the thirsty agent), and the source of suction is the person's mouth and lungs.

The Process of Comprehension

Given this speculative characterization of the knowledge activated by nominal terms, we now consider the role played by these terms in the process of comprehension. We have noted several indications that the vehicle plays the major role in guiding the comprehender toward a resemblance. Schemata in the vehicle domain tend to be the predominant source of constraints by which the topic domain is interpreted. In the tree trunk-straw sentence, for example, the comprehender is invited to apply the straw schemata to the tree trunk domain, that is, to create similar relational systems among appropriate entities in the new domain. In this creative process of schematization, the comprehender will seek to instantiate both the transformational and structural aspects of the vehicle domain: the trunk as the vertical cylindrical space, the leaves and branches as the thirsty agents and source of suction, the earth as the receptacle, ground water as the fluid, the transport of water as the fluid flow, etc. This process will lead to a growth in knowledge when the topic domain is successfully organized by schemata that are unfamiliar or unconventional in that context. The activation of knowledge by topic and vehicle terms is apparently asymmetric: the topic terms activate a comparatively unconstrained system of potential relationships, while the vehicle terms activate specific schemata that are more tightly constrained. Rather than relaxing normative constraints on the topic, the comprehender seeks to impose specific constraints from the vehicle domain, so that the topic term (object) participates in a specific type of event or relationship characteristic of the vehicle. This model of the comprehension process predicts a marked "specificity of encoding" for topic terms, a prediction that is consonant both with our prompted recall data and with the recall of nonmetaphoric materials (for example, Tulving and Thomson, 1973; Bransford and McCarrell, 1974).

At this point we have been able to provide only a rough framework for a model of the comprehension process. More explicit formulations will become possible as solutions are found to several remaining puzzles. One puzzle is how the terms in a metaphoric sentence activate the vehicle domain. The single nominal term straws, for example, clearly underspecifies all of the structures and events in the elaborated vehicle domain. One factor that
shapes the resulting domain is the "familiarity" or "salience" of certain events or relationships in which the object can participate (though this does little more than label the phenomenon). The results for both the arbitrary metaphor and the sentence completion tasks provided circumstantial evidence that vehicles are more likely than topics to be exemplary instances of the grounds, and, conversely, that the grounds are more likely to be salient schemata for vehicles than for topics. Another factor is the use of contextualizers to constrain the comprehender's search for the intended schema. For example, finding the appropriate schema for straws is aided by extending the predicate phrase to are straws for thirsty X. Also of great importance is contextualization of the topic. Topic terms often appear mixed into the predicate, as in thirsty leaves and branches (tree trunks), giraffes of a city (skyscrapers), and warts on the landscape (billboards). These phrases aid in delimiting the appropriate schema and lead listeners to supply comparable entities in the vehicle domain. This was evident in paraphrases like the following (where even the ordering of topic and vehicle was reversed): Giraffes are skyscrapers of the jungle; Giraffes with other animals are like the skyscrapers in the city. Thus, it is not sufficient to argue that the topic is "passively" schematized by salient properties of a vehicle domain: the topic and vehicle terms interact in specifying the ground (see Black, 1962; Verbrugge, 1977).

A second puzzle for future research is to identify the constraints that govern successful schematization. The topic domain does not accept all transformations with equal ease. It is easier to schematize tree trunks as straws than as babies with pacifiers. There must be compatibility constraints operating between the topic and vehicle that govern what relations from the vehicle domain can be extended successfully or easily. These compatibility constraints, defined over abstract relations, may play a major role in judgments about metaphorical force and quality. Attributive conceptual theory has sought to define these constraints in terms of weighted conventional attributes and typically defines grounds as novel attributes transferred to the topic. But simply attaching new labels to a topic term does not provide a basis for determining when the process proceeds easily or successfully. The attributes represented in an attributive concept are properties that an object manifests in a heterogeneous set of conventional events or relationships. We are doubtful that a metric defined over such attribute lists can predict the ease of interpreting the topic in an unconventional event or relationship. Such a prediction may be possible only for transparent and uninformative metaphors (such as the skyscraper-giraffe sentence). We suspect that it will prove easier to define constraints on metaphorical transformations if structural concepts are defined from the outset by potential transformations under which they remain invariant. As we noted above, this may allow theoretical development of a single type of comprehension process that generates interpretations for both metaphoric and literal sentences.

A third major puzzle is how to characterize the topic domain so that it has sufficient functional plasticity to allow for novel schematization, yet is sufficiently constrained that various vehicle domains are differentially compatible with it. Models based on normative associations do not have sufficient plasticity to explain how the topic domain can be schematized in
radically different ways in the context of different vehicles. Associative network models, semantic feature theories, and models of attributive conceptual knowledge all seek to interpret novel sentences by reference to fixed connections established over long experience. Such systems grow only by accretion; radical transformations, contingent on specific contexts, are not normally envisioned or easily modeled. Our results suggest that the topic domain is highly malleable as a function of the vehicle context; a topic is not "recognized" during recall unless the ground specifies the relationship by which it was originally schematized. To accommodate metaphoric growth in a general theory of comprehension, we need to characterize semantic structures by systems of organization that allow for greater functional plasticity than is possible in heterogeneous networks and hierarchies. (See Turvey, Shaw and Mace, in press, for discussion of an analogous problem.)

The Process of Recall

If metaphoric grounds are characterized as abstract relations, their effectiveness as prompts poses a challenge for current models of the recall process. Experimental studies of word and sentence memory have emphasized the identities of the terms encountered during acquisition. It is assumed that these are central to the cognitive representation of the event and serve as the focus for organizational processes and recall. Verification probes, recognition foils, and recall prompts usually contain terms that appeared in the original event or terms "associated" with the acquisition terms in earlier experience. Our results, like those of Tulving and Thomson (1973), suggest that acquisition terms do not have a stable specific identity or set of associations in different contexts of interpretation. A prompting event may "identify" the related acquisition event by means of an abstract transformational resemblance. A relation of nominal or associative identity is not necessary as a basis for reminding.

Thus, the first stage of prompted recall may be the recognition of a recently experienced event (see Jenkins, Wald and Pittenger, in press). If this recognition proceeds on the basis of sufficient resemblance, not of identity, reminding itself can be considered a metaphoric process. The second stage of recall would be a process of regenerating the specific sentence constituents that originally led the comprehender to experience the event. The often regenerative nature of the second stage is evidenced by the kinds of paraphrases we cited above. This proposed model reverses the order of generation and recognition processes found in many two-stage models of recall (for example, Bahrick, 1970; Tulving and Thomson, 1973) and emphasizes the role of abstract relationships, rather than specific elements, as agents in the recognition phase. Considerable research is needed to determine the conditions under which recognition is likely to occur, and to differentiate between direct recognition of the earlier event (as in a déjà vu experience) and recognition mediated by some kind of search process. Subjects reported both types of recognition experience.

It is difficult to determine what kinds of representation, if any, to attribute to the comprehender of a metaphor. In these experiments, the grounds were formulated as verbal predicates. Since these were effective
prompts, it is tempting to assume that they prompted recall by accessing similar representations created during acquisition. This approach would accept the common assumption that sentence meaning is coded internally by means of a predicate or propositional notation system. An alternative possibility is that sentence comprehension is not representationally mediated, but is a vicarious engagement of the processes underlying perception and action (see Werner and Kaplan, 1963; Arnheim, 1969; Gibson, 1971; Verbrugge, 1977). Our characterization of domains in terms of structural and transformational invariants is consistent with this proposed alternative. If the role of a verbal prompt is to allow the listener to re-experience (recognize) a relation experienced at acquisition, prompts specifying that relation in any modality should be effective, that is, the relations may be abstract with respect to medium (verbal, optical, acoustic), as well as specific contents (tree trunks, straws, hoses, pipes). While propositional projections of abstract relations have considerable heuristic value for theoreticians, attributing these representations to the comprehender may preclude successful explanation of plasticity in word use and the imaginal processes that underly comprehension. Further study of the conditions for successful recall of metaphors may help direct the current controversy over "mental representation" (see Pylyshyn, 1973; Shepard, 1975; Kosslyn and Pomerantz, 1977).

The formal proposition has, for too long, been taken as the prototypical linguistic form. It has shaped the way we define the problems of expression, comprehension, and representation. For example, in many psycholinguistic tasks, subjects are asked to judge the validity of propositions about the outside world or about an artificial "experimental world." The subjects usually cooperate by implicitly adopting the experimenter's constraints: they respond realistically, conventionally, and normatively. Little attention is given to the possibility that the propositions rejected as "false" might be valid in appropriate metaphorical contexts. Many linguists and psychologists have adopted a similar implicit standard when developing theories for interpreting "deviant" expressions: they have attempted to normalize such expressions into standard axiomatic form, so that the canons of verification and inference will apply. While these exercises have some value for purposes of traditional linguistic description, they are of doubtful value as a basis for a theory of creativity in language use. The metaphoric "speech act" invites cognitive processes distinct from those engaged in accessing and verifying facts. Metaphor invites pretending, imagining, reasoning by analogy; in its more powerful forms, it requests a perception of resemblances by means of an unconventional reshaping of identities. The study of metaphorical competence in adults challenges us not to limit these processes to the nursery room and the therapist's couch, but to see them as crucial phenomena in the psychology of everyday life.

REFERENCES


Cassirer, E. (1923) *Substance and Function.* (Chicago: Open Court).


Psychologist 9, 660-694.


