

## Interpreting and Evaluating Metaphors

ROGER TOURANGEAU

NORC, Washington, DC 20036

AND

LANCE RIPS

University of Chicago

Although several current theories concern the meaning of metaphors, relatively few empirical studies directly examine that issue. This paper reports a series of three experiments on how subjects interpret and evaluate metaphors. In the first, subjects characterized the meaning of metaphors (e.g., *The eagle is a lion among birds*), as well as their tenors (*eagle*) and vehicles (*lion*). The results suggest that many of the features in the interpretation of the metaphor are emergent and are not well established parts of preexisting conceptions of the tenor or vehicle. Conversely, many features thought to characterize both tenor and vehicle are not seen as relevant to the interpretation. In the second study, subjects rated features listed in the first study. The rating scales included salience, relationality, and distinctiveness—the key variables according to earlier theories of metaphor. These ratings indicate that, although they may help describe the features included in the interpretation of a metaphor, factors such as salience imbalance or relationality do not predict the rated goodness of the metaphors. In the final study, subjects judged which of two types of interpretation better captured the meaning of metaphors drawn from modern poetry. For each metaphor, one interpretation was based on features shared by the tenor and vehicle; the other was based on emergent features. Subjects overwhelmingly preferred the interpretations based on the emergents. © 1991 Academic Press, Inc.

Despite continuing interest in metaphor since Aristotle's initial speculations on the topic, there is relatively little evidence on the central question of how we understand metaphors and other kinds of figurative language. The relative dearth of evidence has not prevented an abundance of theories. At

least three models of metaphor—Gentner's structure-mapping model (Gentner, 1982; Gentner & Clement, 1989), Ortony's salience imbalance model (Ortony, 1979), and Tourangeau and Sternberg's domains interaction theory (1981, 1982; Tourangeau, 1982)—have been the focus of psychological research, and recently Glucksberg and Keysar (1990) proposed a fourth. Still other models are prominent in linguistics and artificial intelligence (Carbonell, 1983; Hobbs, 1983; Lakoff & Johnson, 1980; Lakoff & Turner, 1989). We concentrate on the three models on which the most psychological research has been conducted.

Although these three viewpoints differ about a number of issues, perhaps the most important differences involve the meaning of metaphors. Following terminology first introduced by the literary theorist I. A.

We thank Marshall Abrams, Paul Estin, Lazar Primack, Tim Ryan, and Rita Walter for their help in collecting and coding the data discussed here. We also thank Dedre Gentner, Albert Katz, and Boaz Keysar for their comments on earlier drafts of this paper. Finally, we thank Julie Johnson for her aid in interpreting the metaphors in Study 3. This research was supported in part by NIMH Grant MH39633. Correspondence about the paper reprint requests can be sent to Roger Tourangeau, NORC, 1350 Connecticut Avenue, Suite 500, Washington, DC 20036, or to Lance Rips, Department of Psychology, University of Chicago, 5848 S. University Avenue, Chicago, IL 60637.

Richards (1936), we refer to the principal subject of the metaphor as its topic, or *tenor*. We refer to the secondary subject, which is used to characterize the tenor, as the *vehicle* of the metaphor. The *ground* of the metaphor refers to the features that are central to its interpretation. For example, the metaphor *The eagle is a lion among birds* has eagle as its tenor and lion as its vehicle. Its ground would include such features as stateliness and impressiveness that the metaphor attributes to eagles.<sup>1</sup>

For Ortony (1979), metaphors are statements of nonliteral similarity, and metaphors, like other similarity statements, are understood by finding features, or predicates, that are shared by the tenor and vehicle. With literal statements of similarity, the two subjects share features that are salient for both; with metaphors and similes, however, salient features of the vehicle are shared by, but less salient for, the tenor. According to Ortony, these matching features that are salient for the vehicle but not the tenor are crucial for interpreting a metaphor.

According to Gentner and Clement (1989), interpreting metaphors (and analogies more generally) typically requires "mapping" the structure of the vehicle onto that of the tenor: "An analogy is a mapping of knowledge from one domain (the base) into another (the target) which conveys that a system of relations that holds among the base objects also holds among the target objects." This mapping process involves transferring predicates from the vehicle (i.e., the base) to the tenor (i.e., the target) or matching the predicates of the two. The predicates that define the vehicle's structure—the system of relations

among the base objects—are crucial to the interpretation of a metaphor, and these "relational" predicates should figure prominently in descriptions of the meaning of metaphors.

Finally, according to Tourangeau and Sternberg (1982), the interpretation of metaphors involves mapping features that apply within one domain to an object from a different domain; because a given feature may not apply outside its original domain (at least not literally), tenor and vehicle need not share any features. For example, the metaphor "men are wolves" is not usually intended to mean that men are carnivorous, although that is a feature shared by men and wolves; instead, the metaphor suggests that men are competitive in their dealings with other men, a feature that does *not* characterize wolves. Tourangeau and Sternberg's theory claims that although the metaphor's meaning does depend on the features that distinguish the vehicle from other members of its domain, these distinguishing features may be transformed before they are applied to the tenor, as when the predatory nature of wolves is mapped onto the competitive dealings of men.

At the most basic level, then, current models of metaphor differ over which features, or predicates, of the vehicle determine the meaning of the metaphor. For Ortony, the critical features are salient features of the vehicle that are less salient for the tenor; for Gentner and Clement, the critical ones are shared predicates that convey the "system of relations" in which the vehicle is embedded; for Tourangeau and Sternberg, the features that are crucial for interpreting metaphors are those that distinguish the vehicle from other things of the same type or, more generally, define the vehicle's position relative to a system of concepts.

There is little evidence that tests any of these theories unambiguously or that allows a critical comparison of the three. Most of the data on the comprehension of

<sup>1</sup> In traditional lexicography, a "feature" is a one-place predicate such as +*unmarried* that is part of the definition of a term. Our own use of the term is broader, including any predicate (with any number of arguments) that subjects believe to characterize a term or metaphor.

metaphor concerns the speed with which metaphors are interpreted (e.g., Harris, 1976; Ortony, Schallert, Reynolds, & Antos, 1978). These reaction time studies examine the idea that metaphorical readings for a sentence are attempted only after a literal reading has failed (e.g., Chomsky, 1964; J. Katz, 1964; Searle, 1979; Van Dijk, 1975). If this two-stage comprehension model is correct, metaphors should take longer to interpret than literal sentences; in fact, the evidence suggests that metaphors are understood just as quickly as literal sentences. Other evidence against the two-stage model indicates that metaphorical readings are activated automatically, even when a literal reading for the sentence is possible (Glucksberg, Gildea, & Bookin, 1982; Keysar, 1989).

We are aware of only three papers that directly examine interpretations of nonliteral comparisons (Gentner & Clement, 1989; A. Katz, 1982; Ortony, Vondruska, Foss, & Jones, 1985), but unfortunately the results of these studies are contradictory. Ortony and his coworkers report a series of studies that examined interpretations of similes (e.g., *reality is like a sledgehammer*). The interpretations were either provided by the experimenter (Study 3) or generated by the subjects (Study 4). Ratings data suggested that the features constituting the grounds were more salient for the vehicle than for the tenor of the comparison, as the salience imbalance model predicts. Katz's earlier study had looked at metaphor interpretations and also found support for the salience imbalance model. The results reported by Gentner and Clement, on the other hand, find no evidence for salience imbalance; instead, their studies indicate that a predicate's "relationality" (its importance in defining the relational structure to be mapped onto the tenor) determines whether it is included in the metaphor's interpretation.

Most of the data supporting the domains interaction theory is only indirectly related

to the issue of how we understand metaphors. Based on ratings of the aptness of metaphors, Tourangeau and Sternberg (1981) showed that subjects prefer metaphors in which the tenor can be seen as occupying a position within its category, or domain, that is analogous to the position of the vehicle; further, their evidence suggests that dissimilarities between the categories from which the tenor and vehicle are drawn also contribute to the aptness of the metaphor (see Trick & Katz, 1986, for a replication of these findings).

The data are thus fragmentary at best. Basic questions about the interpretation of metaphors remain unanswered: Does the ground, or interpretation, of a metaphor consist of shared features? Do some features relevant to the interpretation characterize neither the tenor nor the vehicle, but "emerge" from the metaphor? Does the interpretation process differ when the tenor is too new or unfamiliar to be compared to the vehicle?

We carried out several studies to address these questions. In the first study, we compare features that characterize the tenor and vehicle with those listed as part of the interpretation of the metaphor; this study also examines the relations between characteristics of a metaphor's interpretation and judgments about the metaphor, such as its aptness. The aptness ratings are important for two reasons. First, they allow us to compare our findings with those of the earlier studies by Tourangeau and Sternberg. More important, the aptness ratings should help resolve ambiguities in the main findings involving comprehension. For instance, it could be that, although subjects *can* interpret metaphors in terms of features showing salience imbalance, such interpretations lead to low aptness ratings. The ratings data allow us to evaluate such possibilities.

In the second study, we gathered additional ratings of the features that entered into the ground of the metaphors in Exper-

iment 1. These additional ratings included saliency and relationality, and they therefore help test the claims of the various theories. The final study compares two types of interpretations of metaphors—one based on features shared by the tenor and vehicle and the other based on emergent features. This experiment uses literary metaphors from a sample of poems, unlike most previous studies of metaphor (although see Katz, Paivio, & Marschark, 1985, for an exception).

#### EXPERIMENT 1: CHARACTERISTICS OF THE GROUNDS OF METAPHORS

The logic of Study 1 was quite simple. The subjects listed what they knew about a number of objects (e.g., horses, ferries, eagles, and lions). They also listed what each of several metaphors said about their tenors, where the metaphors in question combined two of the objects (e.g., *The ferry is a horse among ships*, *The eagle is a lion among birds*). We compared the object descriptions with the interpretations, classifying each feature in the interpretation according to whether it characterized the tenor of the metaphor, the vehicle, both, or neither. In addition, the subjects rated the aptness and comprehensibility of the metaphors, and we explored which sorts of features contributed most to these qualities of the metaphor.

According to most theories of metaphor, a crucial part of the interpretation process is finding features shared by the tenor and vehicle. We were therefore interested in seeing how subjects would interpret metaphors that involved unfamiliar tenors—tenors with essentially no features to compare to those of the vehicle. For this reason, some of the subjects interpreted and rated metaphors whose tenors were omitted, such as *X is a lion among birds*.

#### Method

This experiment used 12 metaphors of the form “The *tenor* is a *vehicle* among

*domain*” (e.g., *The eagle is a lion among birds*). The metaphors were formed by pairing tenors and vehicles randomly, creating a set of items that varied widely in their aptness. This permitted us to test some earlier claims about what makes a metaphor a good one and to explore whether the interpretations of good metaphors systematically differed from those of bad ones. Table 1 lists the full set of metaphors. Half of the subjects saw the metaphors in their original form; half saw the same items, but with the tenor replaced by a letter (e.g., *X is a lion among birds*). We will refer to the original items as the metaphors with tenors and the others as metaphors with missing or omitted tenors. Subjects wrote down their interpretation of each of the metaphors, evaluated each for goodness, aptness, and comprehensibility, and listed properties for each of the content terms in the metaphor (e.g., eagle, lion, and bird).

*Tasks.* Each subject carried out two

TABLE 1  
MEAN GOODNESS RATINGS OF METAPHORS USED IN  
EXPERIMENTS 1 AND 2

Metaphors	Mean goodness rating
The eagle is a lion among birds	7.85
The ferry is a horse among ships	5.55
Richard Nixon is a goose among U.S. historical figures	5.15
Alexander Hamilton is a battleship among U.S. historical figures	4.50
Ulysses S. Grant is an elephant among U.S. historical figures	4.16
The rowboat is a Herbert Hoover among ships	3.98
The clipper is a pelican among ships	3.25
The heron is a freighter among birds	2.58
The water buffalo is an Abraham Lincoln among land mammals	2.45
The mongoose is an ostrich among land mammals	2.42
The buzzard is a Franklin Delano Roosevelt among birds	2.18
The gorilla is a troop transport among land mammals	2.02

*Note.* Each mean is based on 40 subjects.

tasks—metaphor ratings and object descriptions. The subjects rated separately how apt, good, and comprehensible each metaphor was on an 11-point scale. They then listed phrases describing what the metaphor attributed to its tenor. The crucial part of the instructions read as follows:

After you rate each metaphor on the three scales, please write down what the metaphor says about its subject. For example, with the metaphor "Mozart is the chocolate torte among composers," Mozart is the subject, and you would write down all the things the metaphor says about Mozart. Please try to express each idea in a simple sentence or phrase. If you think the metaphor says that Mozart is artful and elaborate, list both these ideas as separate entries. Be complete; don't be afraid to list things that seem too "obvious" or subjective.

The three rating scales were highly intercorrelated (all correlations above .90).<sup>2</sup> We present the results in terms of the goodness scale. Because of the high intercorrelations, using the other scales would not change any of the conclusions. Table 1 lists the mean goodness ratings for each of the metaphors with a tenor.

The object description task involved separately describing the 12 tenors, 12 vehicles, and the 4 domains (birds, land mammals, ships, and U.S. historical figures) from which they were drawn. Each of the 28 terms appeared on a separate page. The subjects were instructed to list as many things as they could remember about each one and to take as long as they wanted to complete the task. Half of the subjects first rated and interpreted the 12 metaphors and then generated object descriptions; the

metaphors did not accompany the tenors or vehicles in the object description task. The other half completed the 28 object descriptions before seeing the metaphors; they then rated and interpreted the 12 metaphors.

Subjects received the metaphors either with the tenors present or with the tenors omitted. This variable was crossed with task order, giving four experimental groups with 20 subjects each. Regardless of which metaphors they received, all subjects described all 28 objects (including the 12 tenors missing from the metaphors with the omitted tenors).

*Stimuli.* We constructed the metaphors for this experiment by sampling four categories—ships, birds, land mammals, and U.S. historical figures—from a group of eight that Tourangeau and Sternberg (1981) had previously scaled. From each category, we randomly selected six instances from the 20 scaled by Tourangeau and Sternberg, designating three as tenors and three as vehicles. Finally, we formed the set of metaphors in Table 1 by randomly pairing the tenors in one category with a vehicle in each of the other categories.

Each page of the metaphor rating booklet displayed a metaphor at the top. Below it were three 0-to-10 scales with the ends labeled *not apt* and *apt*, *bad* and *good*, and *incomprehensible* and *comprehensible*. Below the three scales was a series of 40 numbered lines on which subjects listed phrases describing what the metaphor attributed to the tenor. In the object description booklets, each page showed the object term (e.g., eagle) at the top followed by numbered lines for the object's properties. The order of the pages in the metaphor and object booklets was randomized for each subject.

*Feature compilation.* Some of the features that subjects listed were synonyms or near-synonyms (e.g., intimidating and threatening) or differed in other minor ways. Treating such items as separate features would create a misleading picture of

<sup>2</sup> We were somewhat surprised at the high correlations of the aptness and goodness ratings with the ratings of comprehensibility. Tourangeau and Sternberg (1981) had found much lower correlations between aptness and comprehensibility ratings, and, more recently, Gerrig and Healy (1983) presented evidence that the evaluation and comprehension of a metaphor can be dissociated. Within the small set of metaphors here, however, aptness and comprehensibility were empirically indistinguishable.

the results, understating the degree of overlap in the responses. For this reason, we counted two items as instances of the same feature if they met any of five criteria: (a) the items were synonymous (*fierce* and *ferocious*), (b) they were alternative spellings or abbreviations (e.g., *vice pres* and *vice president*), (c) they differed only in degree (usually because of an intensive modifier—e.g., *very strong* and *strong*), (d) they shared a key predicate (e.g., *climbs trees* and *climbs*), or (e) they differed only in level of abstraction (e.g., *water bird* and *bird*). These rules—especially the one for combining synonyms—are hardly very precise. In general, we tried to combine features that were very close in meaning, but not merely related. For example, the feature “is a leader” was not combined with “king of the animals.” In reporting the number of features for the metaphors in what follows, we will mean the number of features after merging in this way.

The feature lists were a melange of different types of predicates. For example, some 64 features were listed by at least two subjects as characterizing lions. The most popular were that lions live in Africa (mentioned by 36 subjects), that they are cats (16 subjects), that they are gold in color (16 subjects), that they are large (42 subjects) and strong (15 subjects), that they are the kings of beasts (47 subjects), that they have manes (54 subjects), and that they are predators (70 subjects). Some 37 features were listed at least twice for eagles. Those mentioned most often were that one type of eagle is the bald eagle (mentioned by 21 subjects), that eagles are an endangered species (19 subjects), that they have feathers (9 subjects), that they fly (19 subjects), that they are large (7 subjects), and that they are predators (32 subjects). Eight features were included by two or more subjects as part of the meaning of *The eagle is a lion among birds* but were not listed as characterizing either lions or eagles. The most popular of these was that the metaphor implied that eagles were *respected*.

**Subjects.** Eighty University of Chicago undergraduates took part in the study. They were recruited by means of an advertisement in the student newspaper, which requested subjects for experiments on memory and reasoning. All were native speakers of English. The subjects were paid \$5.00 for their time and were run individually or in groups.

### Results

The main results of the first study bear on two questions: What is the source of features that are included in interpretations of the metaphors? How do different types of features (e.g., those shared by tenor and vehicle) contribute to a metaphor's goodness?

**Metaphor interpretations.** We compiled a master list of the features listed for each object and each metaphor, as described earlier. We then classified each feature that was listed as part of the interpretation, or ground, of the metaphor according to whether it was also mentioned for the tenor or the vehicle. Table 2 presents the proportion of features in the ground that were unique to the ground, shared with the tenor of the metaphor (but not the vehicle), shared with the vehicle (but not the tenor), and shared by both tenor and vehicle. The table is based only on the subjects who received the metaphors with tenors. The results are presented in two different ways—

TABLE 2  
FEATURES IN THE INTERPRETATION OF THE  
METAPHORS BY SOURCE

Types (%)	Tokens (%)	Source
58	24	Unique to the interpretation
26	15	Also listed for vehicle but not for the tenor
8	39	Also listed for the tenor but not for the vehicle
8	22	Also listed for both tenor and vehicle

*Note.* The results are based on 40 subjects and 12 metaphors.

with each feature counted once regardless of the number of subjects who listed it as part of their interpretation ("types") and with each feature counted as many times as it was mentioned by different subjects ("tokens"). As examples of the kinds of features that fit into these categories, take the metaphor *The eagle is a lion among birds*. For this item, the feature *is respected* was unique to the ground, *is mean* was part of the ground and the vehicle, *is a symbol* is part of the ground and the tenor, and *is a predator* is part of the ground, vehicle, and tenor.

It is obvious from Table 2 that only a minority of the features that are included in the interpretation of a metaphor are shared by both the metaphor's tenor and vehicle. The converse is also true—only a minority of the features (45.7% of the types) shared by the tenor and vehicle are included in the interpretations of the metaphors. For instance, the feature *has eyes* was listed for both eagle and lion but not for the interpretation of *An eagle is a lion among birds*. A feature's being shared by the tenor and vehicle is thus neither a necessary nor sufficient condition for its being incorporated into the interpretation of a metaphor.

We analyzed the number of features listed for each metaphor by the four feature categories in Table 2, treating metaphor as a random factor. Leaving aside for the moment the data from the groups that received the metaphors with the tenors omitted, we found highly significant differences across categories:  $F(3,11) = 37.5$  ( $p < .001$ ) if types are analyzed;  $F(3,11) = 2.93$  ( $p < .05$ ) if tokens are analyzed. These category differences were similar whether subjects had rated the metaphors first or listed features first; there were no significant interactions between category and task order. In the types analysis, pairwise comparisons of the category means (by the Tukey method) revealed that subjects listed significantly more features as part of the interpretation only (an average of 5.2 per metaphor) than features shared with the vehicle (2.3); in ad-

dition, subjects listed significantly more features in both of these categories than features shared by the interpretation and tenor (0.8) or by the interpretation and both tenor and vehicle (0.7). In the tokens analysis, none of the pairwise differences is significant.

*Relation to goodness.* The more features listed in the interpretation, the better the metaphor. For the metaphors with tenors, the correlation between the total number of features in the interpretation and the mean rating of the metaphor's goodness is .82; the correlation is slightly lower (.74) if tokens are counted instead of types. Both correlations are highly significant ( $p < .01$ ).

Table 3 examines the relation between goodness ratings and the number of features listed from different sources; the correlations are again based on the data from the group that received metaphors with specified tenors. Only the correlations involving features shared by tenor, vehicle, and interpretation are significantly greater than zero, although the correlation between the number of features unique to the interpretation and the mean goodness ( $r = .51$ ) is marginally significant ( $p < .10$ ) in the tokens analysis. We also note that features that are shared by the tenor and vehicle but are *not* included in the interpretation do not seem to detract from the metaphor (see the last row in the table).

*Metaphors with omitted tenors.* Accord-

TABLE 3  
CORRELATION OF NUMBER OF FEATURES AND MEAN  
GOODNESS BY SOURCE

Types	Tokens	Source
.20	.51	Unique to the interpretation
.33	.35	Also listed for vehicle but not tenor
.19	.04	Also listed for tenor but not vehicle
.63	.64	Also listed for both tenor and vehicle
.32	.42	Listed for both tenor and vehicle but not interpretation

Note. Each correlation is across the 12 metaphors.

ing to any theory in which shared features are central to the interpretation process, metaphors which omit their tenors present problems; at a minimum, such metaphors will require a comprehension process quite different from the one for metaphors with explicit and familiar tenors. The metaphors with omitted tenors may be interpreted more readily (since the tenors place fewer constraints on the interpretation) or with greater difficulty (since they share few, if any, features with the vehicle). Either way the comprehension process should be different.

The ratings data suggest that the metaphors with the missing tenors are seen as similar in goodness (mean goodness rating of 4.4) and comprehensibility (mean comprehensibility of 5.1) to the metaphors with tenors (means of 3.9 for goodness and 4.9 for comprehensibility). Analyses of variance indicate that the effect of including the tenor is not significant for either variable.

The number of features listed in the interpretation for each metaphor tends to be somewhat higher when the metaphor had a tenor (10.2) than when the tenor was omitted (8.9), but the difference is not significant. If we consider just those features unique to the interpretation and those common to the interpretation and the vehicle, then the main effect of specifying the tenor is significant: The metaphors with missing tenors had significantly more unique features (an average of 6.2 types) and features derived from the vehicle in the interpretations (4.1 types) than the metaphors with tenors (5.2 unique types and 2.3 types from the vehicle);  $F(1,33) = 17.3$  for the overall effect of specifying the tenor,  $p < .01$ . The results are the same when we examine tokens— $F(1,33) = 19.8$ ,  $p < .01$ . However, this does not imply that, in the absence of a specific tenor, subjects transfer features wholesale from the vehicle to the tenor. Once again, the majority (60%) of the feature types listed in the interpretation of metaphors with omitted tenors were fea-

tures that were not included in descriptions of the vehicle. It is likely that this is because the tenor's domain (e.g., *birds* in the case of *X is a lion among birds*) places constraints on what features can be included in the interpretation.

### Discussion

Perhaps the key finding of this initial study is that the interpretation of a metaphor includes "emergent" features, features not ordinarily seen as characterizing either the tenor or vehicle. It may be that the large number of emergent features indicates that the subjects had a hard time interpreting the metaphors and that their interpretations included a number of wild guesses. If so, one would expect such features to detract from the rated quality of the metaphor; but in fact, Table 3 indicates that the number of emergent features is positively related to ratings of the goodness of the metaphor.

Almost as striking is the fact that features shared by the tenor and vehicle do not dominate the interpretations (comprising only 8% of the types and 22% of the tokens). In addition, many of the features that are common to the tenor and vehicle do not figure in the ground.

These results place limits on previous theories of metaphor. Saliency imbalance, for example, cannot be a complete theory of metaphor interpretation if many of the features making up the interpretation are parts of neither tenor nor vehicle. The saliency imbalance hypothesis predicts that features in the interpretation should have high vehicle saliency and low tenor saliency. It is certainly possible that the saliency was so low that subjects simply neglected to list these features for the tenor. However, it is much more difficult to understand why more of the features in the interpretation were not listed for the vehicle. Since the saliency imbalance hypothesis predicts that the interpretative features are salient for the vehicle, subjects should



have mentioned them. Yet Table 2 shows that the majority of features in the interpretation (whether types or tokens) did not appear in the vehicle listings.

Of course, this does not imply that salience imbalance plays no role in processing metaphors. The data are consistent with the possibility that salience imbalance does characterize the subset of features that are common to the tenor, vehicle, and ground. Since this subset appears to be especially important in determining the metaphors' goodness, the effect of salience imbalance may lie primarily in the way metaphors are evaluated. The data, thus, cannot completely rule out the salience imbalance hypothesis. Similarly, they cannot rule out Gentner's structure mapping model either; for instance, Gentner and Clement may be right that features in the interpretation are relational ones and that relational features enhance a metaphor's goodness.

Another issue left unresolved by our initial study is the nature of the emergents. Despite the positive correlation of the number of emergent features with the mean rated goodness, it is still possible that the emergents are irrelevant to the interpretation and represent features subjects generate as they tried to cope with basically incomprehensible metaphors. We undertook a second study to examine these unresolved issues.

#### EXPERIMENT 2: SALIENCE, RELATIONALITY, AND DISTINCTIVENESS

In our second study, a new group of subjects rated the features generated in Experiment 1. The new ratings included the salience and distinctiveness of the features for the tenor and vehicle, the degree that they specified relations rather than simple attributes, and their relevance to the interpretation of the metaphor. Our aim in collecting these data was to examine the nature of the emergent features more closely (e.g., are they truly relevant to interpreting a metaphor?) and to test the divergent pre-

dictions of the various models about which features are likely to be part of the interpretation and which contribute to its goodness.

#### *Method*

To select the features for this study we examined the features generated in Experiment 1 by subjects who had seen the metaphors with tenors. We chose from this set all features that were either a part of the interpretation of a metaphor or a common feature of the tenor and vehicle, dropping those that were unique to the tenor or unique to the vehicle.

*Ratings.* For each feature, subjects rated on 10-point scales its salience for the tenor and for the vehicle, its distinctiveness for the tenor and vehicle, its relationality, and its relevance to the interpretation of the metaphor. For each of these sets of ratings, the respondent received a booklet containing the instructions for the rating task, the features to be rated, and the rating scales.

The booklet for the salience ratings consisted of 24 pages, one for each of the tenors and vehicles used in Experiment 1. Each page listed one of the 24 objects followed by a number of features. We instructed the subjects to rate each feature's salience with respect to the object, and to help clarify this, we asked them to bear three criteria in mind—the feature's general applicability for the object in question, its centrality to the object, and the degree that it was characteristic of the object. (These criteria correspond to the separate measures used by Ortony et al., 1985.) A feature was to be rated high in salience if it was high on any of these three criteria.

The booklet for the distinctiveness ratings followed a similar format—one of the tenors or vehicles was printed at the top of the page, followed by a list of features to be rated. In making the distinctiveness ratings, subjects were told to rate the extent that the feature served to distinguish the object in question from other things of the same type.

The booklets for the relationality ratings gave one of the metaphors, followed by a list of features to be rated. (The metaphor was included to clarify the meanings of the features.) Subjects were instructed to rate the degree that the feature involved a relation rather than a simple attribute (cf. Gentner & Clement, 1989). The instructions gave *hit* as an example of a relation ("since it related one object (a hitter) to another (the thing hit)") and *red* as an example of an attribute ("since it is a characteristic of a single object").

The final ratings booklet followed the same format as the booklet for the relationality ratings—each page was headed by one of the 12 metaphors and listed a number of features. This booklet asked subjects to rate whether the feature in question was relevant to the interpretation of the metaphor.

Half of the subjects received the distinctiveness and salience booklets in the first part of the experimental session and the relationality and relevance booklets in the second part. For the remainder, we reversed this order. Within this constraint, we employed all eight possible orders of presentation of the booklets, assigning two subjects to each order. The pages of the booklets were randomly permuted for each subject.

**Subjects.** We recruited 16 subjects from the same pool as in Experiment 1. None

had taken part in the first study. They were paid \$5 for participating.

### *Results and Discussion*

We organize our discussion of the results from each set of ratings around two questions: (a) How do the different categories of features (e.g., emergents) differ in salience, relationality, and so on? and (b) Which feature characteristics are associated with good metaphors? Table 4 summarizes the evidence on the first question by displaying the mean ratings on our measures for each feature category.

**Salience.** Columns 1 and 2 of Table 4 present the average salience ratings with respect to the tenor and vehicle. These data support the validity of the feature listings from the first study. In general, the features that subjects produced in describing a given term in Experiment 1 are also rated more salient for that term in the present study: The mean salience rating with respect to the tenor was 6.7 for features that had been listed for the tenor (rows 3, 4, and 5 of Table 4) but 4.3 for features that had not (rows 1 and 2). Similarly, mean salience with respect to the vehicle was 7.3 for features listed for the vehicle in Experiment 1 (rows 2, 4, and 5), but 5.0 for those that had not (rows 1 and 3). These differences yield a reliable interaction between the source of the feature and the term being rated (tenor or vehicle),  $F(1,60) = 116.3, p < .001$ . As

TABLE 4  
MEAN SALIENCE, DISTINCTIVENESS, RELATIONALITY AND RELEVANCE RATINGS BY SOURCE OF FEATURES

Source	Salience		Distinctiveness		Relationality	Relevance
	Tenor	Vehicle	Tenor	Vehicle		
Unique to interpretation	4.73	5.79	4.50	5.66	5.50	5.27
Also listed for vehicle, but not tenor	3.80	7.65	3.59	7.49	5.18	5.20
Also listed for tenor, but not vehicle	7.00	4.22	7.12	3.94	4.91	5.03
Also listed for both tenor and vehicle	7.74	8.01	7.57	7.82	5.12	7.33
Listed for tenor and vehicle, but not interpretation	5.47	6.25	5.17	5.89	3.79	4.18

*Note.* Means are based on 16 subjects and 12 metaphors.

we expected, the emergent features have relatively low salience for both the tenor and vehicle.

According to Ortony's (1979) salience imbalance hypothesis, features included in the interpretation of a metaphor should be more salient for the vehicle than for the tenor. Consistent with this view, we found that the average salience rating for all features included in the interpretation (first four rows of Table 4) was higher for the vehicle (6.4) than for the tenor (5.8),  $F(1,60) = 30.0$ ,  $p < .01$ . As can be seen from Table 4, however, salience imbalance was not particularly marked for the subset of features that were both shared by the tenor and vehicle and included in the interpretation (those in the fourth row of the table); for this group of features, the mean salience was 7.7 for tenor and 8.0 for the vehicle— $F(1,60) = 1.5$ ,  $p > .10$ . Moreover, salience imbalance is also present for shared features that were *not* included in the interpretation (mean salience of 5.5 for the tenor and 6.2 for the vehicle,  $F(1,60) = 12.7$ ,  $p < .01$ ).<sup>3</sup>

The salience imbalance model does not correctly predict the rated goodness of the metaphors. On the basis of the model, one would expect better metaphors to have in-

terpretations with features that are especially salient for the vehicle and especially nonsalient for the tenor. In fact, the mean salience with respect to the tenor is positively correlated with goodness ( $r(10) = .83$ ;  $p < .01$ ), and the mean salience for the vehicle is uncorrelated with goodness ( $r(10) = -.02$ ). Thus, salience imbalance—as measured by the difference between the salience for the vehicle and salience for the tenor of the features included in the interpretation—is *negatively* correlated with goodness ( $r(10) = -.64$ ;  $p < .05$ ), just the reverse of the predicted effect.

*Relationality.* According to Gentner's structure mapping model (Gentner, 1982; Gentner & Clement, 1989), the predicates that are crucial to the interpretation of a metaphor tend to be those that specify relations rather than simple attributes. In line with this view, the mean relationality ratings for features included in the interpretation (5.2) was significantly higher than the ratings of features not included in the interpretation but listed as part of both tenor and vehicle (3.8),  $F(1,60) = 25.5$ ,  $p < .01$ .

There was, however, little variation across metaphors in the relationality of the features listed in their interpretations. As a result, there was no significant correlation ( $r(10) = .17$ ,  $p > .10$ ) with the mean rating of the goodness of the metaphor.

*Distinctiveness.* According to Tourangeau and Sternberg's domains interaction theory (1981, 1982; Tourangeau, 1982), features that distinguish the vehicle from other things of the same type (the vehicle's within-domain features) are the basis for interpreting a metaphor. This suggests that the features in the ground will be more distinctive for the vehicle than for the tenor. This appears to be the case—features in the interpretation are on average higher in distinctiveness for the vehicle (6.2) than for the tenor (5.7),  $F(1,60) = 25.8$ ,  $p < .01$ .

However, like salience imbalance and relationality, distinctiveness of the vehicle's features did not predict goodness ratings. For the features included in the interpreta-

<sup>3</sup> The finding that the features included in the interpretation are more salient for vehicle than tenor may appear to conflict with the results of Experiment 1. In that study, we found that relatively few of the features in the interpretation were also mentioned for the vehicle, a fact we took as evidence against the salience imbalance theory. We note, however, that the Study 2 result concerns the *relative* salience of features for tenor and vehicle. It may be that these features are somewhat more salient for the vehicle than for the tenor, even though they do not really describe either very well. This is consistent with the finding in Study 1 (see Table 2, column 1) that an even smaller proportion of the features included in the interpretation are shared with the tenor than with the vehicle. The point is that salience imbalance offers an incomplete theory of metaphor interpretation because it fails to explain where the features in the interpretation come from. This gap remains even if these features exhibit salience imbalance.

tion, average distinctiveness for the vehicle correlated with goodness a nonsignificant .04. By contrast, the distinctiveness of the interpretative features for the tenor is highly correlated with goodness,  $r(10) = .85$ ,  $p < .01$ . This pattern of correlations is similar to what we have noted for the salience ratings; both the salience and distinctiveness of features for the tenor, but not the vehicle, correlate with goodness ratings. This parallel is a natural one, since we might expect salience and distinctiveness to be highly intercorrelated, and indeed they are— $r(10) = .98$  for both the tenor and the vehicle.

**Relevance.** The relevance ratings allow us to check whether features that subjects in Experiment 1 listed as part of the interpretation are in fact relevant to that interpretation. The last column in Table 4 confirms this relationship. The average rating for features in the interpretation (the first four rows of Table 4) is 5.7, whereas the average rating is 4.2 for features common to the tenor and vehicle but not part of the interpretation (the last row);  $F(1,60) = 52.4$ ,  $p < .01$ . We note that the relevance ratings are also consistent with the findings of the previous study in highlighting the importance of features shared by tenor, vehicle, and the interpretation. In Experiment 1, the number of features in this category correlated highly with ratings of the goodness of the metaphor (see Table 3); in Experiment 2, features in this same class were rated as particularly relevant to the interpretation of the metaphor (Table 4).

These ratings also confirm that the emergent features are relevant to the metaphor's interpretation. Subjects rated these features as more relevant to the interpretation than those shared features that were left out of the interpretation,  $F(1,60) = 16.7$ ,  $p < .01$ ; they also rated them as at least as relevant as every other class of features listed in the interpretation, except those shared by tenor, vehicle, and interpretation.

We would expect that the greater the relevance of the features listed in the interpre-

tation the better the metaphor should be. In accordance with this, there was a .56 correlation between rated relevance and rated goodness,  $p = .06$ . Thus, of the various measures in Table 4, a metaphor's goodness seems to depend on the salience or distinctiveness for the tenor of the features listed in its interpretation and marginally on the relevance of these features for the interpretation; but goodness apparently does *not* depend on the salience or distinctiveness of the interpretative features for the vehicle, on salience imbalance, or on whether the interpretative features are attributes or relations.

#### EXPERIMENT 3: EMERGENT VERSUS SHARED FEATURES

The results of Experiment 2 confirmed the main finding of Experiment 1: Interpretations of metaphors may include emergent features that are not particularly characteristic of either the tenor or vehicle. Although consistent with the traditional notion that metaphors can produce novelty, the emergents cast doubt on the idea that interpretations of metaphor are built on features shared by the tenor and vehicle.

Despite the converging results of the first two studies, there is still room for doubt about the importance of the emergents. Both of the studies are based on the same set of 12 metaphors. These metaphors, being somewhat artificial, may lack any good interpretation. A related point is that the metaphors were all of a certain form ("The *tenor* is a *vehicle* among *domain*") and that form may encourage a type of interpretation that is peculiar to this class of metaphors. It is possible that, had the metaphors permitted them, subjects would have preferred interpretations based on shared features to those based on emergents.

In our third study, we attempted to resolve these doubts. We selected a new set of metaphors, a set that was reasonably representative of metaphors as they appear in modern poetry. These metaphors often have an open character that permits multi-

ple interpretations; however, it cannot be said that they lack any plausible interpretation at all. For each metaphor in this set we constructed two interpretations. One consisted of features that, in our judgment, were shared by the tenor and vehicle and that were more salient for the vehicle than for the tenor; the other consisted of emergent features. The subjects rated the salience of the features in the interpretations (these ratings serve as a check on our classification of the interpretative features as shared or emergent); the subjects also rated the adequacy of each interpretation. This study thus directly pits the two types of interpretation against each other.

#### Method

We selected a set of 15 literary metaphors and similes from a standard anthology of modern poetry. For each one, we made up two interpretations—one based on features shared by the tenor and vehicle and the other based on emergent features. Most of the metaphors we chose were extended metaphors, with multiple points of metaphoric comparison. In such cases, our interpretations focus on one of these aspects. For example, one of the metaphors (actually, in this case, a simile) was *Like a bear to its floe, I clambered to bed*. One of the interpretations for this metaphor was based on features shared by ice floes and beds—*The bed is flat, smooth, and pale in color*. The other consisted of an emergent feature—*The bed provides relief from the environment*.

The subjects had two tasks. They rated the adequacy of the metaphor interpretations; they also rated the salience for the tenor and vehicle of the features comprising each interpretation. For instance, they rated the salience for beds of the property *flat, smooth, and pale in color*. One group of subjects completed the salience task before reading the metaphors and rating their interpretations; the other group carried out the two tasks in the opposite order.

*Ratings.* For the interpretation task, the

subjects received a 15-page ratings booklet. Each page of the booklet contained a metaphor followed by two brief interpretations. One of the interpretations was based on emergent features; the other was based on features shared by the tenor and vehicle. The subjects' task was to read the metaphor and to judge how adequately the interpretations expressed the meaning of the metaphor. The instructions asked the subjects to rate each interpretation according to whether it captured "part of the interpretation of the metaphor"; subjects made their ratings on a 10-point rating scale that followed each interpretation. One end of the scale was labeled *definitely not part*; subjects were told to use values at this end for "statements that definitely do not capture the meaning of the metaphor." The other end of the scale was labeled *definitely part*; the subjects were told to use values at this end for statements that did "capture the meaning of the metaphor." The Appendix lists the 15 metaphors and the two interpretations of each one.

There were two versions of the interpretations rating booklet. In one version, the interpretation based on emergents was presented first for eight randomly determined metaphors and second for the other eight. The other booklet reversed the order in which the two interpretations appeared. The order of the 16 metaphors was randomly determined for each subject.

The subjects' other task was to rate the salience for the tenor and vehicle of the features on which the interpretations were based. The instructions for this task were essentially the same as those used for the salience ratings in Experiment 2. The subjects made their ratings in a booklet containing 60 items. Each item consisted of an object (the tenor or vehicle of one of the 15 metaphors), a feature, and a 10-point scale for rating the salience of the feature for the object. Altogether there were 60 items, formed by pairing each tenor and vehicle with the emergent or shared features that comprised one of the interpretations of the

relevant metaphor. The order of the 60 items was randomly determined for each subject.

*Stimuli.* We selected a simple random sample of 57 poems from the *Norton Anthology of Modern Poetry* (edited by Richard Ellman), a widely used compilation of works in English by 20th century poets and certain of their 19th century precursors (e.g., Emily Dickinson). Thirty-six of the poems had at least one metaphor or simile; we selected the first one from each poem. We then culled out those that were difficult to interpret when taken out of context.

We had an outside reader, a high school English teacher, write interpretations for each of the 16 remaining metaphors. The interpretations presented to the subjects were derived from these (see the Appendix). One interpretation was couched in terms of features that, in our judgment, literally characterized both tenor and vehicle. The other consisted of features not ordinarily characteristic of either the tenor or vehicle.

As a final check on the interpretations, we asked a group of 12 subjects to rate each of them for "poeticness" and "vividness." Half of the subjects first rated how poetic they thought the interpretations were, using a scale from 0 ("not poetic") to 10 ("extremely poetic"). They then rated the vividness of the same interpretations on a similar scale (0 = "not vivid," 10 = "extremely vivid"). The remaining subjects made the vividness ratings before the poeticness ratings. None of the subjects saw the actual metaphors, and they received the interpretations in random order.

The means from these preliminary ratings showed that one of the metaphors had an emergent interpretation rated much more poetic than its shared feature interpretation. We therefore eliminated this metaphor from the main experiment. The other 15 metaphors had interpretations that were quite comparable in poeticness and vividness. The mean poeticness ratings for this set were 4.5 for the emergent interpre-

tations and 4.3 for the interpretations based on shared features ( $SE = 0.60$ ). Similarly, the mean vividness of the emergent-based interpretations was 4.4 and the mean vividness of the interpretations based on shared features was 4.6 ( $SE = 0.68$ ). The two types of interpretation were also about the same length (6.8 words on average for the emergents and 8.3 words for the shared features).

*Subjects.* We recruited 16 subjects from the same pool as those in Experiments 1 and 2. None had taken part in either of the earlier studies or in the preliminary rating study. They were paid \$5 for participating.

### *Results and Discussion*

We carried out three main lines of analysis. The first examined the salience ratings as a function of task order (salience ratings before or after the interpretation ratings), the object rated (tenor or vehicle), and the type of feature (emergent or shared). The second examined the interpretation ratings as a function of task order and type of interpretation (based on emergent or shared features). The final analysis looked at the relationship between the salience ratings (in particular, salience imbalance) and the ratings of the adequacy of the interpretations.

*Salience ratings.* The salience ratings revealed that subjects regarded the shared feature interpretations as more salient for the vehicles than for the tenors of the metaphors; however, they viewed the emergent interpretations as more salient for the tenors than the vehicles. For the shared features, the mean salience was 5.6 for the tenors and 6.7 for the vehicles; for the emergents, the mean salience was 6.5 for the tenors and 5.5 for the vehicles. This reversal produced a significant interaction between type of interpretation and part of the metaphor,  $\text{quasi-}F(1,21) = 6.36, p < .05$ . Thus, the relative salience of the shared feature interpretations is in accord with the salience imbalance theory, whereas the salience of the interpretations based on emergents conflicts with that theory. If the sa-

lience imbalance view is correct, the shared feature interpretations should fit the metaphors better than the emergent interpretations.

Overall, the salience of the shared features was about the same as the salience of the emergents ( $M = 6.2$  vs.  $6.0$ ). However, the difference widened for those subjects who rated salience before seeing the metaphors. For this group, the mean ratings were  $6.1$  for shared and  $5.5$  for emergent features. The group who completed the salience ratings after viewing the metaphors produced means of  $6.2$  for the shared features and  $6.5$  for the emergents. The interaction between group and type of interpretation was significant when metaphors were treated as the random factor ( $F(1,14) = 5.15, p < .05$ ) and was marginally significant when subjects were treated as random ( $F(1,14) = 3.10, p < .05$ ). Apparently, the metaphors tended to increase the salience of the emergent properties.

*Interpretation ratings.* The main finding from the analysis of the ratings of the interpretations was the subjects' overwhelming preference for the interpretations based on emergent rather than shared features. The emergent-based interpretations received an average rating of  $8.6$ , while those based on shared features received an average rating of  $3.7$ ; quasi- $F(1,24) = 100.8, p < .001$ . The difference was in the same direction for all 15 metaphors.

*Salience imbalance and interpretation ratings.* Our final analysis examined the relation between the salience imbalance of an interpretation (i.e., the difference between the mean salience rating for the vehicle and the mean salience rating for the tenor) and its adequacy as an interpretation of the metaphor (i.e., its mean rating in the interpretation rating task). If the salience imbalance hypothesis is right, these two variables should be positively correlated, indicating that more adequate interpretations consist of features that are more salient for the vehicle than for the tenor. There is a nonsignificant positive correlation ( $r = .08$ ) for

the interpretations based on shared features; however, for the interpretations based on emergents, the correlation is both negative and marginally significant— $r(13) = -.44, p < .10$ .

The results thus demonstrate that interpretations based on emergent features may be preferred to ones based on shared features. Depending on the type of interpretation, salience imbalance was either negligibly associated with ratings of the adequacy of the interpretation or negatively associated.

### GENERAL DISCUSSION

The first study showed that shared features, which characterize both the tenor and vehicle, may not dominate the interpretation of a metaphor, but that emergent features, which do not characterize either, can play a central role. That study also showed that subjects prefer metaphors with detailed interpretations (as measured by the number of features). The second study reinforced these results by demonstrating that subjects rate the emergent features as relevant to the interpretation of the metaphor. The results from Experiment 2 also underscored the importance of an appropriate interpretation: When the features in the interpretation actually apply to the tenor (i.e., are highly salient and distinctive for the tenor), then subjects rate the metaphor as a good one. Neither salience imbalance nor relationality, however, was related to the goodness of the metaphors in Experiment 2. The final study further demonstrates the importance of emergents to the interpretation of metaphors. For the metaphors in Experiment 3, we were able to construct interpretations consisting of emergents that were clearly preferred to interpretations consisting of shared features.

Two earlier results foreshadow our findings concerning the importance of emergent features. The first concerns the ground as a cue for recalling a metaphor. Verbrugge and McCarrell (1977) demonstrated that the ground could be an effective recall cue for the metaphor, even when, among subjects

who had not seen the metaphor, it was not an effective cue for either the tenor or vehicle. Based on their results, Verbrugge and McCarrell argued that the ground of a metaphor consists not of shared features but of a new view of the tenor, based on a "novel schematization of the topic domain" (p. 494). The second finding involved data from a lexical decision task. Camac and Glucksberg (1984) showed that lexical decisions were faster for pairs of words that were associatively related than for pairs formed randomly; however, pairs consisting of the tenor and vehicle of a metaphor were no faster than the random pairs. Once again, this finding suggests that the ground of a metaphor is not based on existing associations between the tenor and vehicle.

#### *Implications for Theories of Metaphor*

Ortony's salience imbalance view and Gentner's structure mapping model both single out specific subsets of shared features as having a crucial role in the interpretation of metaphors. Although Glucksberg and Keysar's (1990) theory of metaphor is new and its implications for the comprehension process are not yet entirely clear, it too seems to share the assumption that metaphors are understood in terms of shared features. According to Glucksberg and Keysar, a metaphor asserts that the tenor belongs to the class exemplified by the vehicle; this class is, however, defined by "a set of properties that are attributable to both" tenor and vehicle—that is, by the features they share (Glucksberg & Keysar, 1990, p. 11). All three of our studies suggest that these models may overstate the importance of shared features. Further, the data from Experiment 2 suggest that neither salience imbalance nor "relationality" serves as the principle determining which metaphors are regarded as good ones, and the data from Experiment 3 indicate that the salience imbalance of the features in an interpretation may be negatively related to its adequacy.

Nevertheless, there are some hints that imbalance and relationality may both play positive roles in interpreting metaphors. We found in Experiment 2 that features of the interpretation are generally more salient for the vehicle than for the tenor and that the same features are more relational than those shared features left out of the interpretation. One simple explanation of these findings might be that, although salience imbalance and relationality cannot account for the whole interpretation, they at least account for the piece of it that overlapping features supply. According to this hypothesis, when subjects decide which shared features belong in the interpretation, they tend to choose those that are relational or that exhibit salience imbalance. For example, given a metaphor like our standard *The eagle is a lion among birds*, people will ignore simple properties that are equally characteristic of tenor and vehicle (*has eyes, is animate*) and concentrate instead on "lion-like" properties that are less obvious aspects of eagles. Such properties could be ones that are especially salient for lions (*is strong*), relate lions to other kinds of things (*is large*), or both (*is a predator*). Our findings on emergent features show that shared properties are not the only source for the interpretation. However, the shared features are seen as particularly relevant to the interpretation and do contribute strongly to ratings of the metaphor's goodness. For these reasons, salience and relationality would still be critical factors.

Although this explanation may be correct for relationality, it runs into trouble in the case of salience imbalance. Table 4 shows that salience imbalance is actually larger for those common features of tenor and vehicle that are not part of the interpretation than for those that are (see rows 4 and 5). If salience imbalance determines which shared features are part of the interpretation, then one would expect a difference in the opposite direction. Getting around this problem seems to require complicating the notion of salience imbalance. For example,



the data in Table 4 suggest that the shared features that made it into the interpretation are more salient overall than ones that were left out. We might therefore want to amend the notion of salience imbalance to take into account, not only the difference between a feature's salience for the vehicle and tenor, but also the absolute level of salience for the vehicle. The idea is that features with low absolute levels of salience for the vehicle will not be included in the interpretation, no matter what their salience for the tenor might be. Since this suggestion is obviously ad hoc, we do not pursue it here; however, it represents one way to resuscitate the salience imbalance hypothesis if we are willing to countenance some additional complications.

The structure mapping model and the domains interaction view of Tourangeau and Sternberg share the assumption—first advanced by Max Black (1962, 1979)—that metaphors involve projecting a structure from one domain onto a second domain. Lakoff and his colleagues (Lakoff & Johnson, 1980; Lakoff & Turner, 1989) have gone even further to argue that such mappings across domains can become conventionalized. This emphasis on a mapping from one domain to another receives support from developmental research by Keil (1986), indicating that once children can understand one metaphor linking two domains they are also able to understand other metaphors involving these domains. In a similar vein, Lehrer (1978) has found that multiple metaphors mapping one domain onto another tend to enter the language at the same time.

How does the notion of a mapping from one domain to another help explain the findings of our three studies? First, it is clear that these mappings often require the transformation rather than the transfer of features that apply to the vehicle. Lakoff and Turner, for example, argue that several specific metaphors rest on a basic underlying metaphor in which things that are greater in some way are mapped onto those

that are higher (e.g., costly goods are said to have high prices); in the parlance of Lakoff and Turner, the basic metaphor asserts that MORE IS UP. Of course, expense will not necessarily characterize objects that are physically high up. Similarly, Tourangeau (1982) showed that a metaphor about a fictional politician ("Donald Leavis is the George Wallace of Northern Ireland") evoked the inference that Leavis was anti-Catholic; but anti-Catholicism, though it corresponds to Wallace's racism, is not itself a characteristic of Wallace. Having been projected onto the tenor, the features central to the ground may no longer apply to the vehicle.

If the interpretations of metaphors are built on features that are based on features of the vehicle but that do not themselves characterize the vehicle, then it is also possible that they will not necessarily jibe with our existing views about the tenor either. There are several ways such discrepancies can arise. First of all, the metaphor may convey a new conception of the tenor, or, as Verbrugge and McCarrell (1977) put it, "a novel schematization of the topic domain." As an example, consider the metaphor *marriage is a zero-sum game* (discussed by Black, 1979); we can readily interpret this metaphor even though it requires us to see marriage in a new, rather unconventional way. "Interaction" accounts of metaphor, particularly those of Black (1962) and Beardsley (1958), have long emphasized the changes in perspective produced through metaphor. Such reconceptualizations will be based on features that are not ordinarily associated with the tenor.

The features of the ground may not match existing features of the tenor for more everyday reasons. One is that the tenor is new or unfamiliar and the metaphor is adding to our knowledge of it. Like literal statements, metaphors can be used to characterize new subjects, with which we were previously unacquainted. Many of the literary metaphors used in Experiment 3 are of

this type. Most of us, for example, are in no position to know whether the speaker in Adrienne Rich's "Necessities of Life" really used herself like a dry bulb (see the first item in the Appendix). This information is not part of our conception of the speaker until the metaphor makes it one; in such cases the interpretation of the metaphor will consist largely of features that are *not* already salient for the tenor. The use of metaphors to convey new information about a topic is probably quite common; consider how easily the subjects in Experiment 1 were able to interpret the metaphors with missing tenors. A final reason why the interpretative features may not be existing features of the tenor is that the metaphor says something that conflicts with our existing view. The picture of marriage conveyed by *marriage is a zero-sum game* may strike many of us not only as novel but false. Despite our disagreement with this metaphor (that is, the mismatch between the features in the ground and the features of the tenor), we can still understand it. Nevertheless, the correlations of Experiment 2 suggest that subjects prefer metaphors that attribute properties that are salient or distinctive of their tenors, that actually fit the tenor. If people already have knowledge of a tenor, then they will prefer metaphors that agree with this knowledge to ones that contradict it. (If agreement with the metaphor affects the evaluation of a metaphor but not its interpretation, this would constitute another example of the dissociation between the two processes; cf. Gerrig & Healy, 1983.)

Whenever a metaphor presents a new picture of a familiar topic or one that clashes with an existing picture, or when the metaphor adds to our developing view of a new topic, emergent features are likely to figure prominently in the interpretation. In each of these situations, the features of the ground will consist not of relatively inaccessible features of the tenor, but rather of true emergents, features predicated of the tenor for the first time.

### *Is Metaphor Comprehension Special?*

The view that emergents are crucial to the comprehension of metaphors may suggest that metaphor comprehension is somehow special, involving processes that are not needed for the interpretation of literal language. There are, however, several good reasons to reject this view. First, as we noted earlier, the data from reaction time studies indicate that metaphors are understood just as quickly as literal sentences; this seems to rule out the idea that they require additional processes, beyond the ones needed for literal sentences. Second, the linguistic evidence cited by Lakoff and his colleagues suggests that there is a continuum between novel metaphors, metaphors based on conventional mappings, "dead" metaphors, and literal language. It would be surprising if categorically different processes were required to interpret these different types of language; it seems unlikely, for example, that there is much difference in the comprehension process for *prices rose* and *prices increased*. (See also Hobbs, 1983, who argues that the same sorts of inferential processes are need to interpret metaphorical and literal sentences.)

But the most compelling reason to reject the argument that emergents imply a unique process of metaphor comprehension is the finding that emergents may also be important in interpreting certain literal constructions. Recent papers by Murphy (1988) and by Medin and Shoben (1988) demonstrate that emergents may be crucial in understanding complex concepts conveyed by noun phrases. For example, an *apartment dog*, according to Murphy's results, has features that are not particularly characteristic of either apartments or dogs in general. Similarly, subjects inferred that an *empty store* is one that is losing money—something that is not especially typical of either stores or empty things. What is especially intriguing about Murphy's results is how closely they parallel our own:

Shared features are not always important in the definition of complex concepts, whereas emergent features can be crucial. Thus, whether phrases are literal or metaphorical, comprehension is likely to involve more complicated inferential processes than finding features shared by two concepts.

Murphy interprets his results in terms of conceptual schemata; he argues that the adjective (e.g., *empty*) is used as the basis for filling in a slot in the schema for the noun concept (*store*). Similar views have been proposed within the literature on metaphor (see, for example, Lakoff & Turner, 1989). It may be, however, that an account of emergent properties in metaphor requires knowledge structures that are less fixed than schemata, perhaps like those comprehenders sometimes seem to construct on the fly (Medin & Shoben, 1988; Rips, 1989). Though we regard these approaches as promising, it is still too soon to say whether they will bear fruit for our understanding of figurative language.

#### APPENDIX: METAPHORS AND INTERPRETATIONS USED IN EXPERIMENT 3

- (1) Scaly as a dry bulb thrown into a cellar I used myself, let nothing use me.—Adrienne Rich, "Necessities of Life."  
The poet is enriched by the world's organic material. (Shared feature)  
The poet is consumed by using her own experience. (Emergent)
- (2) Nobody thought, to judge by the cut of him, he could ever be anything other than normal—silent and straight as the shaft of a shovel stuck for the night at the edge of a field.—Desmond O'Grady, "The Father."  
The person is spindly and of proper height. (Shared feature)  
The person is solid and morally upright. (Emergent)
- (3) A light he was to no one but himself.—Robert Frost, "An Old Man's Winter Night."  
The man produces heat and radiance. (Shared feature)  
The man is isolated and solitary. (Emergent)
- (4) Now as I was young and easy under the apple boughs about the lilting house and happy as the grass was green.—Dylan Thomas, "Fern Hill."

The happiness of the child is dependent on nourishing substance. (Shared feature)

The happiness of a child is essential and intrinsic. (Emergent)

- (5) Your knees are a southern breeze—or a gust of snow.—William Carlos Williams, "Portrait of a Lady."  
Knees are in a constant state of agitation and motion. (Shared feature)  
Knees are pleasantly surprising and thrilling. (Emergent)
- (6) We have gone out in boats upon the sea at night, lost, and the vast waters close traps of fear around us.—Robert Duncan, "Passage over Water."  
Night waters are nearly invisible, restricting movement. (Shared feature)  
Night waters are enclosing and surrounding. (Emergent)
- (7) The path, winding like silver, trickles on, bordered and even invaded by thinnest moss.—Edward Thomas, "The Path."  
The path is found in or near the ground, surrounded by dust. (Shared feature)  
The path is narrow and bright, surrounded by darkness. (Emergent)
- (8) What, then, was war? . . . an infection of the common sky that sagged ominously on the earth.—Robert Graves, "Recalling War."  
War has large effects that ensue from small causes. (Shared feature)  
War is an all-encompassing negative condition. (Emergent)
- (9) The spirit of man is a bird in a cage that beats on the bars with a goodly rage and fain would be free.—Hugh McDiarmid, "I Heard Christ Sing."  
The spirit of man resides in a contained vessel. (Shared feature)  
The spirit of man fights against restriction. (Emergent)
- (10) Like a bear to its floe, I clambered to bed.—Randall Jarrell, "90 North."  
The bed is flat, smooth, and pale in color. (Shared feature)  
The bed provides relief from the environment. (Emergent)
- (11) To shift, to fetch, to drive, to shed, to pen are acts I recognize, with all they mean of shepherding the unruly, for a kind of controlled woolgathering is my work too.—Cecil Day Lewis, "Sheepdog Trials in Hyde Park."  
Writing is an occupation offering little compensation. (Shared feature)  
Writing consists of imposing order on difficult ideas. (Emergent)
- (12) I desire a name for you, nice, as a right glove fits.—Ann Spencer, "At the Carnival."  
A name is often one of a pair. (Shared feature)  
A name is appropriate. (Emergent)

- (13) By some derision of wild circumstance . . . last night we fell together to achieve a light eclipse of years.—Edwin Arlington Robinson, "Reunion."  
The reunion at night is surrounded by darkness. (Shared feature)  
The reunion at night blocks out time. (Emergent)
- (14) Had I the heavens' embroidered cloths, enwrought with golden and silver light, the blue and the dim and the dark night and light and the half light, I would spread the cloths under your feet.—William Butler Yeats, "Aedh Wishes for the Cloths of Heaven."  
The heavens form a protective covering. (Shared feature)  
The heavens are rich, varied, beautiful. (Emergent)
- (15) Let her lift the oceans in her eyelids to the moon.—Louis Simpson, "The Green Shepherd."  
Her eyes act as vessels for water. (Shared feature)  
Her eyes are soulful. (Emergent)

## REFERENCES

- BEARDSLEY, M. (1958). *Aesthetics*. New York: Harcourt, Brace, & World.
- BLACK, M. (1962). *Models and metaphors*. Ithaca, NY: Cornell University Press.
- BLACK, M. (1979). More about metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 19–43). Cambridge: Cambridge University Press.
- CAMAC, M., & GLUCKSBERG, S. (1984). Metaphors do not use associations between concepts, they are used to create them. *Journal of Psycholinguistic Research*, 13, 443–455.
- CARBONELL, J. (1983). Derivational analogy in problem solving and knowledge acquisition. In R. S. Michalski (Ed.), *Proceedings of the International Machine Learning Workshop* (pp. 12–18). Urbana, IL: Department of Computer Science, University of Illinois at Urbana-Champaign.
- CHOMSKY, N. (1964). Degrees of grammaticality. In J. Fodor & J. Katz (Eds.), *The structure of language*. Englewood Cliffs, NJ: Prentice-Hall.
- GENTNER, D. (1982). Are scientific analogies metaphors? In D. Miall (Ed.), *Metaphor: Problems and perspectives*. Brighton: Harvester Press.
- GENTNER, D., & CLEMENT, C. (1989). Evidence for relational selectivity in interpreting analogy and metaphor. In G. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press.
- GERRIG, R., & HEALY, A. (1983). Dual processes in metaphor understanding: Comprehension and appreciation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 667–675.
- GLUCKSBERG, S., & KEYSAR, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, 97, 3–18.
- GLUCKSBERG, S., GILDEA, P., & BOOKIN, H. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, 21, 85–98.
- HARRIS, R. (1976). Comprehension of metaphor: A test of a two-stage processing model. *Bulletin of the Psychonomic Society*, 8, 321–324.
- HOBBS, J. R. (1983). Metaphor interpretation as selective inferencing: Cognitive processes in understanding metaphor (Part 1). *Empirical Studies of the Arts*, 1, 17–33.
- KATZ, A., PAIVIO, A., & MARSCHARK, M. (1985). Poetic comparisons: Psychological dimensions of metaphoric processing. *Journal of Psycholinguistic Research*, 14, 365–383.
- KATZ, A. (1982). Metaphoric relationships: The role of feature saliency. *Journal of Psycholinguistic Research*, 11, 283–296.
- KATZ, J. (1964). Semi-sentences. In J. Fodor & J. Katz (Eds.), *The structure of language*. Englewood Cliffs, NJ: Prentice-Hall.
- KEIL, F. (1986). Conceptual domains and the acquisition of metaphor. *Cognitive Development*, 1, 73–96.
- KEYSAR, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language*, 28, 375–385.
- LAKOFF, G., & JOHNSON, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.
- LAKOFF, G., & TURNER, M. (1989). *More than cool reason: A field guide to poetic metaphor*. Chicago: University of Chicago Press.
- LEHRER, A. (1978). Structures of the lexicon and transfer of meaning. *Lingua*, 45, 95–123.
- MEDIN, D., & SHOBEN, E. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, 20, 158–190.
- MURPHY, G. (1988). Comprehending complex concepts. *Cognitive Science*, 12, 529–562.
- ORTONY, A. (1979). Beyond literal similarity. *Psychological Review*, 86, 161–180.
- ORTONY, A., SCHALLERT, D., REYNOLDS, R., & ANTOS, S. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior*, 17, 465–478.
- ORTONY, A., VONDRUSKA, R., FOSS, M., & JONES, L. (1985). Saliency, similes, and the asymmetry of similarity. *Journal of Memory and Language*, 24, 569–594.
- RICHARDS, I. A. (1936). *The philosophy of rhetoric*. Oxford, UK: Oxford University Press.

- RIPS, L. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge, UK: Cambridge University Press.
- SEARLE, J. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and thought*. Cambridge, UK: Cambridge University Press.
- TOURANGEAU, R., & STERNBERG, R. (1981). Aptness in metaphor. *Cognitive Psychology*, 13, 27–55.
- TOURANGEAU, R., & STERNBERG, R. (1982). Understanding and appreciating metaphors. *Cognition*, 11, 203–244.
- TOURANGEAU, R. (1982). Metaphor and cognitive structure. In D. Miall (Ed.), *Metaphor: Problems and perspectives*. Brighton: Harvester Press.
- TRICK, L., & KATZ, A. (1986). The domain interaction approach to metaphor processing: Relating individual differences and metaphor characteristics. *Metaphor and Symbolic Activity*, 1, 185–213.
- VAN DIJK, T. (1975). Formal semantics of metaphorical discourse. *Poetics*, 4, 173–198.
- VERBRUGGE, R., & MCCARRELL, N. (1977). Metaphoric comprehension: Studies in reminding and resembling. *Cognitive Psychology*, 9, 494–533.

(Received June 25, 1990)

(Revision received December 23, 1990)