Gentner, D., & Clement, C. (1988). Evidence for relational selectivity in the interpretation of analogy and metaphor. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol. 22, pp. 307-358). New York: Academic Press.

EVIDENCE FOR RELATIONAL SELECTIVITY IN THE INTERPRETATION OF ANALOGY AND METAPHOR

Dedre Gentner
Catherine Clement

I. Introduction

Analogies and metaphors are pervasive in language and thought. They range from scientific analogies, such as "electricity is like flowing water," to expressive comparisons such as "the moon, the massy pearl of night," to whole systems of extended meanings, such as "rising and falling GNP" (Lakoff & Johnson, 1980; Kittay & Lehrer, 1981; Nagy, 1974). In this article we ask how such comparisons are interpreted: that is, how the meaning of an analogy or metaphor is derived from the meanings of its terms. We compare four accounts of the interpretation process: Tourangeau and Sternberg's (1981) multidimensional space account, Ortony's (1979) salience imbalance theory, Holyoak's (1985) pragmatic account, and Gentner's (1980, 1983) structure-mapping theory. We argue, both on theoretical and empirical grounds, that despite its representational complexity, only the structure-mapping account is adequate to describe the phenomenona.

Before laying out the theories let us be clear about the issues on which we are focusing and the kinds of comparisons we want to explain. Our concern is with the *interpretation* of metaphor and analogy, that is, with the way in which the meaning of a metaphor or analogy is derived from the prior representations of its constituent terms. We are not aiming at this stage to describe the real-time processing steps by which the inter-

TABLE I

A SELECTION OF COMPARISONS

1. Lemonade is like water.
2. Heat is like water: it flows down a temperature gradient.
3. If we do not plant knowledge when we are young, it will give us no shade when we are old.
4. Sharp wits, like sharp knives, do often cut their owner's fingers.
5. She allowed life to waste like a tap left running. (Virginia Wolfe)
6. I have ventured, /Like little wanton boys that swim on bladders, /This many summers in a sea of glory; /But far beyond my depth. My high-blown pride /At length broke under me; and now has left me, /Weary and old with service, to the mercy /Of a rude stream, that must for ever hide me. (William Shakespeare)
7. The glorious lamp of heaven, the sun. (Robert Herrick)
8. Coffee is like a solar system.

pretation is performed. Rather, the goal is a descriptive model of how the derived meaning of the comparison relates to the initial representations of its terms. Our interest is in what Marr (1982) called the *computational level* rather than in the *algorithmic level*. Or, to put it in Palmer and Kimchi's (1985) terms, we are primarily interested in the informational constraints on interpretation, rather than in the behavioral constraints.

Table I gives an idea of the range of comparisons we must consider. These assertions differ in the degree and kind of similarity they convey. Statement 1 expresses literal similarity: it tells us that most of what we know about water can be applied to lemonade. Statements 2–7 are nonliteral similarity comparisons—either analogies or metaphors. Many of the comparisons could be labeled as either analogy or metaphor (or simile),[1] and for many purposes these two categories are alike. We will combine analogy and metaphor for now; later we discuss their differences. Statement 8 is an *anomaly* because the two terms have nothing in common. It is included to underscore a simple point. In defining metaphor it is not sufficient to differentiate it from literal similarity; we must also differentiate it from anomaly. This means that in order to lay out the interpretation rules for metaphor we must consider what makes a nonliteral comparison apt.

We begin by reviewing three theoretical approaches to analogy and metaphor in increasing order of the complexity of their representational assumptions: (1) the multidimensional space models of Rumelhart and Abrahamson (1973) and Tourangeau and Sternberg (1981); (2) Ortony's salience imbalance theory (Ortony, 1979; Ortony, Vondruska, Foss, & Jones, 1985); and (3) Gentner's (1980, 1983, 1986a, 1988a) structure-mapping theory. In the Section III we present three experiments contrasting structure-mapping with salience imbalance. In the Section IV we take up the question of higher-order knowledge structures in analogy interpretation. We contrast structure-mapping with Holyoak's (1985) pragmatic account, an alternative view which, like structure-mapping, utilizes a complex representational format.

[1]Similes differ from metaphors in that they contain an explicit comparative term such as "like" or "as." However, since available evidence suggests that the underlying interpretation processes for similes and metaphors are highly similar (Reynolds & Ortony, 1980), they will be considered together throughout this article.

II. Three Accounts of Metaphor and Analogy

A. MULTIDIMENSIONAL SPACE MODELS

Rumelhart and Abrahamson (1973) developed a model of analogy based on multidimensional space models of similarity (e.g., Shepard, 1974; Krumhansl, 1978). The model is based on the notion of constructing parallel vectors in a multidimensional space. An analogy such as "Horse:zebra::dog:_____" can be solved by constructing a vector from horse to zebra and then constructing the parallel vector from dog and reading off its end point (which might be "fox" in this case). Tourangeau and Sternberg (1981) extended this model to metaphor. In this model, a metaphor, such as "Brezhnev is a hawk," is a mapping from a base subspace (e.g., birds) to a target subspace (e.g., political figures). A metaphor is understood by constructing a vector from the origin in the target subspace that is parallel to the original vector in the base subspace. The ideal comparison concept is found at the terminus of the target vector; the distance between it and the actual target term is a measure of the within-space fit of the metaphor. In the Tourangeau and Sternberg formulation, the *aptness* of a metaphor is greater the lower the within-space distance and the greater the between-space distance between the base and target subspaces. Tourangeau and Sternberg (1981) found some support for the theory, although chiefly for the within-space predictions. They compared subjects' aptness ratings for metaphors with the within-space and between-space distances obtained from similarity ratings on the items and found the predicted negative correlation between aptness and within-space distance. The predicted positive correlation between aptness and between-space distance was not as strong, although in one study they did find a correlation between quality (the sum of aptness, goodness, interestingness, and likability) and between-space distance. However, a more fundamental difficulty with the multidimensional-space representational format is that it imposes severe limits on the kinds of metaphors that can be modeled. It can capture dimensional relations such as LARGER THAN (hawk,

dove) and can metaphorically match different dimensions: e.g., LARGER THAN with MORE IMPORTANT THAN. But there is no good way to represent event relations, such as PURSUE (hawk, dove), and still less causal relations between events. Given these limitations, the utility of the multidimensional space approach is fairly limited. Indeed, in a later article, Tourangeau and Sternberg (1982) consider other representations and, in particular, remark that semantic networks may be a useful format for representing metaphors involving actions or whole sentences.

B. Salience Imbalance

The next approach we consider is Ortony's (1979) influential salience imbalance model of metaphor. Like Tversky's (1977) contrast model of similarity on which it is based, it uses featural representations rather than mental-distance representations.[2] In Tversky's model, the similarity between two items is a weighted function of their common features less the difference sets of nonshared features, with nonshared features of the target weighted more. Ortony (1979) proposed an extension to Tversky's model in which the salience of a feature is defined relative to the particular object of which it is an attribute. Thus, the same feature can have different salience in two different objects. He suggests that the difference between metaphor and literal similarity is largely due to a difference in the relative salience of the features shared between the base and target. In a metaphorical comparison, such as "billboards are like warts," the shared features (such as *ugly*) are of high salience in the base (warts) and of low salience in the target (billboards). In a literal similarity statement, such as "billboards are like placards," the shared features are of high salience in both the target and the base domain. He suggests that "the imbalance $I(a,b)$ in salience levels of matching attributes of the two terms is a principle source of metaphoricity" (Ortony, 1979, p. 164).

One line of support for the salience imbalance account is the observation that metaphors tend to be strongly directional. Reversing the terms produces a relatively different meaning. For example, the simile "billboards are like warts" conveys something like "billboards are ugly bumps on the landscape." But the reverse order, "warts are like billboards," conveys something like "prominent advertising." In contrast, reversing the terms in a literal similarity comparison produces relatively little change in meaning, as in the statements "billboards are like placards" and "placards are like billboards." Ortony interprets this strong directionality in metaphor in terms of salience imbalance. Since the meaning of a metaphor depends on a match of high-salient features of the base (the second term) with low-salient features of the target (the first term), reversing the order of terms will, in general, change the meaning.

[2]The "features" can refer to relations between objects as well as to simple object attributes.

Ortony's core observation that metaphors tend to display directionality is extremely persuasive. We might ask, then, whether salience imbalance theory could provide an account of how analogies and metaphors are interpreted. To this end, we consider three proposals suggested by Ortony (1979). The first proposal, the one of most interest here, is that salience imbalance provides an interpretation rule for metaphors that specifies how the meaning of a metaphor is derived from the meaning of its terms. Thus, it provides the informational constraint that metaphor interpretation consists of high-salient features in the representation of the base and low-salient features in the representation of the target. The second proposal is that salience imbalance is constitutive of metaphor. That is, the degree of salience imbalance determines the degree to which we take a comparison to be metaphorical. The third proposal is that salience imbalance acts as a real-time processing rule for metaphor. Since there is no evidence for or against this proposal, and since our interest is in informational constraints rather than processing algorithms, we will not be concerned with this possibility further.

Three studies by Ortony et al. (1985) provide general support for the role of salience imbalance in metaphors but do not clearly support the position that salience imbalance constrains the interpretation of metaphors. In the first study subjects were presented with pairs of forward and reverse similes and literal comparisons. They judged which direction was preferable and rated the similarity of the base and target. Order preferences were greatest for similes (suggesting that directionality affected the meaning of similes more than the meaning of literal comparisons), and the differences between similarity ratings of forward and reverse comparisons were strongest for similes. Thus, the results showed stronger directional asymmetry for metaphorical comparisons than for literal statements.

The second and third experiments directly assessed the salience imbalance of comparison statements. In the second study subjects were given propositions taken from interpretations of similes and literal comparisons written by the experimenters.[3] Subjects rated these propositions for their salience with respect to either the base terms or the target terms. Three converging measures of salience were used: applicability, conceptual centrality, and characteristicness. Results revealed higher salience ratings for propositions rated with respect to base terms than for those rated with respect to target terms. Salience imbalance was found for both similes and literal comparisons, but the effects appeared to be stronger for similes.

[3]These interpretations were validated in a previous experiment, in which subjects rated the adequacy of the interpretations. Results suggested that subjects were satisfied with the interpretations.

A third experiment provided similar results. Subjects read either similes or literal similarity statements. They were then given either the base or the target term and asked to provide an attribute of the term that contributed to the metaphor. Finally, subjects rated the salience (actually, the "distinctiveness") of this attribute with respect to the term. (Subjects were told that a distinctive attribute was easily brought to mind, that it was very characteristic of the object, and that it distinguished the object from other objects). Results indicated that the difference in salience of attributes contributed by base terms and target terms was greater for similes than for literal similarity statements.

Although the two latter studies suggest that metaphors show more salience imbalance than literal similarity comparisons, they do not address the specific proposal of interest here—that salience imbalance functions as an interpretation rule. In the second study, subjects rated interpretations provided by the experimenter rather than generating them themselves. In the third study, subjects generated partial interpretations (i.e., they wrote out those attributes of one of the terms that contributed to the metaphor). However, they rated the salience of these attributes with respect to constituent terms *after* they read and interpreted the metaphor. This order opens the way for influence from the metaphor to the object descriptions since placing terms in a similarity or metaphor comparison may increase the subjective salience of their common attributes (Elio & Anderson, 1981; Forbus & Gentner, 1986; Gick & Holyoak, 1983).[4] Rips and Tourangeau have obtained evidence that placing object terms in metaphors does indeed affect the subjective salience of their attributes (L. J. Rips, personal communication, November, 1987). Thus, we cannot assume that the salience of base and object attributes measured after the metaphor has been processed is representative of their normal salience. For this reason, the Ortony *et al.* studies do not address whether or not salience imbalance provides an intepretation rule for metaphor. We will return to this point in the following studies.

C. Structure-Mapping

Gentner's (1980, 1982, 1983, 1988a) structure-mapping theory is aimed at characterizing analogy and differentiating it from ordinary literal similarity. Like featural approaches, the structure-mapping approach is componential, but it assumes a propositional representation in which there are structurally different kinds of components which play different roles in the interpretation process. The basic intuition is that 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

[4]Ortony (1979) suggested that "attribute promotion" may occur in the target; that is, features in the target may be heightened by the metaphor.

also holds among the target objects. Thus, an analogy is a way of noticing relational commonalties independently of the objects in which those relations are embedded. According to this view, in interpreting an analogy people seek a common relational structure. Computationally, the interpretation of analogy requires finding a one-to-one correspondence between the objects of the base and the objects of the target so as to obtain the maximal structurally consistent match in relational structure (Falkenhainer, Forbus, & Gentner, 1986, 1988; Gentner, Falkenhainer & Skorstad, 1987).

In addition to the general structural constraints of one-to-one correspondence and structural consistency, structure-mapping postulates two specific informational constraints on the interpretation of an analogy from its constituent terms. First, what is important is common relations, not common object descriptions. The corresponding objects in the base and target don't have to resemble each other; object correspondences are determined by roles in the matching relational structures. Second, the choice of which relations to match is guided by the principle of *systematicity*: rather than mapping isolated predicates, people prefer to match and carry over systems of predicates governed by higher-order constraining relations. The systematicity principle is a structural expression of peoples' tacit preference for coherence and deductive power in interpreting an analogy.

To take a familiar example, in the Rutherford analogy between the solar system and the hydrogen atom, the intended interpretation consists of a set of common relations: that the nucleus is more massive than the electron (just as the sun is more massive than the planet), that the nucleus attracts the electron, that this plus the mass relation causes the electron to revolve around the nucleus, and so on. Object descriptions are disregarded; there is no attempt to match the nucleus with the sun in color, size, or temperature. Moreover, the choice of the common relational structure of a central-force system is determined by the fact that it is the maximal systematic structure that can be found (or postulated) in both domains. Besides analogy, other kinds of similarity can be distinguished in this framework according to whether the match is one of relational structure, object descriptions, or both. *Analogies* discard object descriptions and preserve relational structure. *Mere-appearance* matches preserve object attributes and discard relational structure. *Literal similarity* matches preserve both relational structure and object descriptions. Figure 1 shows the similarity space formed by varying the degree to which the base and target share different kinds of features.

1. *Representational Distinctions*

We use a propositional representation in which (1) nodes or constants represent concepts treated as wholes and (2) predicates, when applied to the nodes, express propositions about the concepts (cf. Collins & Quillian,

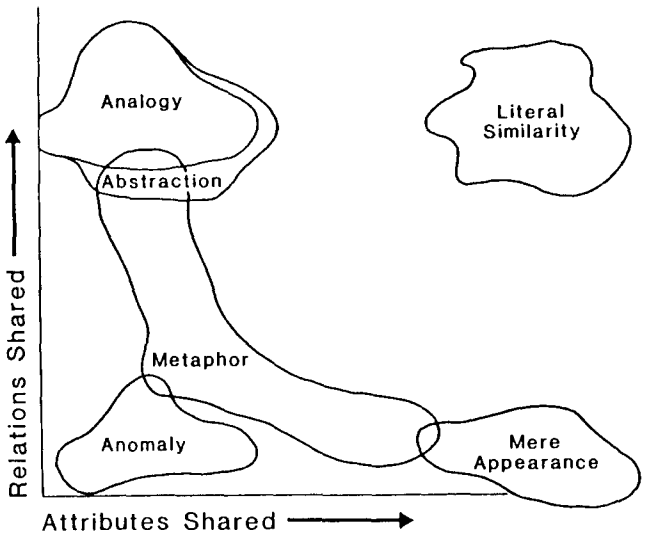

Fig. 1. Similarity space formed by varying the kinds of features shared by the base and target of a comparison.

1969; Miller & Johnson-Laird, 1976; Norman & Rumelhart, 1975; Palmer, 1978; Rumelhart & Ortony, 1977; Schank & Abelson, 1977). Two structural distinctions are important. First, to capture the distinction between object descriptions and rational structure, we distinguish object attributes—predicates taking one argument, such as YELLOW (x)—from relations—predicates taking two or more arguments, such as COLLIDE (x,y).[5] (See Section V for more details.) The second structural distinction is the *order* of a predicate, defined as follows: (1) constants and functions on constants have order 0 and (2) the order of a predicate is 1 + the maximum order of its arguments. Thus, a first-order predicate is one whose arguments are objects. A second-order predicate is one for which at least one argument is a first-order predicate, and so on. For example, if COLLIDE (x,y) and STRIKE (y,z) are first-order predicates, CAUSE [COLLIDE (x,y), STRIKE (y,z)] is a second-order predicate.

It is important to note that these distinctions among predicate types are

[5]Computing the meaning of a one-place predicate may involve an implicit extradomain comparison (Palmer, 1978; Rips & Turnbull, 1980). For example, to comprehend *large* (sun) requires an implicit comparison between the sun and other stars (since *large* for a star is a different size from *large* for, say, a mouse). However, despite this complexity, once the value of the attribute is computed, *large* (sun) can be psychologically treated as a one-place predicate in the domain. In contrast, a predicate such as *larger than* (sun, planet) is inherently a relation between two objects in the discourse. Thus the distinction between one-place and n-place predicates is well formed if all the objects are in the domain of discourse.

intended to apply to *psychological* representations. Logically, the same proposition can be expressed in many formally equivalent ways. For example, a relation $R(a,b,c)$ can be represented perfectly well as a one-place predicate $Q(x)$, where $Q(x)$ is defined to be true just in case $R(a,b,c,)$ is true. But our interest is not in all of the ways a domain could logically be represented, but in how it is psychologically represented at a given time for a given person. The claim is that, given the person's current representations of the base and target, the structure-mapping rules describe the informational constraints on the interpretation of an analogy.

2. *Analogy and Metaphor*

Structure mapping was developed as an account of explanatory analogies, such as those used in science (e.g., Gentner, 1983; Collins & Gentner, 1983, 1987). However, we suggest that a large class of metaphors can also be encompassed in this framework (Gentner, 1982; Gentner *et al.*, 1987). As shown in Fig. 1, we can divide metaphors into categories of relational metaphors, attributional metaphors, and combinations of these.[6] Relational metaphors convey common relational structure and can be analyzed like analogies. Attributional metaphors are mere-appearance matches: their focus is on common object attributes. There are also metaphors that are combinations of attributes and relations. We can exemplify these distinctions with the comparisons in Table I. As discussed previously, Statement 1 is a literal similarity comparison and Statement 8 is an anomaly. Statement 2 would be called analogy, and Statements 3–6 are relational metaphors. (Note that they could easily be described as analogies.) Finally, Statement 7 is an attributional metaphor.

Although metaphors can be either relational or attributional, relational metaphors appear more characteristic of adult metaphorical language. We suggest that people seek relational interpretations of metaphors and prefer metaphors for which such interpretations can be found. Adding this preference assumption results in two general claims. First, in deriving the interpretation of a metaphor from the prior representations of the base and target, people should try to preserve common relations and disregard attributes. Second, people should judge aptness by the degree to which they are successful in finding a relational interpretation.

As a psychological model, structure-mapping is rather elaborate. It assumes that the comprehension of metaphors involves processing of com-

[6]It must be noted that there are metaphors which do not fit the framework, e.g., metaphors that lack structural consistency (see Gentner 1982). An example is Dylan Thomas's "On a star of faith pure as the drifting bread,/As the food and flames of the snow. . . ." Such metaphors are characterized by many cross-weaving connections with no one best mapping between base predicates and target predicates. We will not consider such metaphors here.

plex representational structures and that the matching process is sensitive to distinctions about predicate structure. In contrast, salience imbalance makes far fewer representational assumptions. It requires only a set of features (or predicates) ordered by salience; no structural distinctions are required. It is reasonable to ask whether the elaborate assumptions of structure mapping are necessary, or whether the simpler representational assumptions of salience imbalance are sufficient to account for metaphor interpretation. Therefore, we now present experiments that contrast salience imbalance and structure mapping

III. Experiments Contrasting Structure-Mapping and Salience Imbalance

Structure-mapping and salience imbalance make different predictions about how people derive the meaning of a metaphor from the prior representations of its terms. In this section we describe a series of studies that test these predictions. The method was straightforward. Subjects first wrote out descriptions of individual terms. Then they wrote out interpretations of forward or reverse metaphors containing these terms and rated the metaphors for metaphoricity and aptness. (They were not told about the metaphor task until after they had completed the description task.) Subsequently, judges evaluated the responses according to the predictions of salience imbalance and structure-mapping, as described shortly.

The structure-mapping hypothesis states that people seek interpretations of metaphors that preserve relations from the base and drop object attributes. This generates three specific predictions. First, the metaphor interpretations should include more relations than object attributes. (The assessment of attributionality and relationality is described later). Second, this difference between the amount of relational information and the amount of attributional information should be greater for metaphor interpretations than object descriptions. Third, the more relations subjects can map from base to target, the more apt they should find the metaphor. Therefore, the aptness ratings for metaphors should be positively correlated with the degree to which the metaphor interpretation is relational. No such prediction holds for attributes: there should be either no correlation or a negative correlation between aptness ratings of metaphors and the attributionality of the interpretations.

In drawing predictions from the salience imbalance theory, one difficulty is the lack of a clear definition of "salience." In this study, following a suggestion by Ortony (1979), we operationalized salience as the order of mention of propositions in subjects' object descriptions. With this proviso, salience imbalance makes three predictions. First, if salience imbalance

governs the interpretation of a metaphor, then the chief determinant of which aspects of the object descriptions are used in the metaphors should be salience imbalance; that is the metaphor interpretations should contain a preponderance of features[7] that are mentioned early in the base description and late, if at all, in the target description. The remaining two predictions derive from the claim that metaphoricity depends on salience imbalance.[8] The second is that the rated metaphoricity of forward metaphors should be greater than that of reverse metaphors. This is because the feature matches for the forward metaphors—e.g., "cigarettes are like time bombs"—should satisfy salience imbalance to a greater degree than should the reversed metaphors—e.g., "time bombs are like cigarettes." Third, if the metaphors vary in the degree to which they display salience imbalance, the rated metaphoricity should depend on the degree of salience imbalance: that is, on the degree to which the predicates that appear in the metaphor interpretations appear early in the base descriptions and late in the target descriptions.

A. *Experiment 1*

1. *Method*

a. Subjects. Undergraduate college students (20) from the Cambridge, Massachusetts area, served in the basic metaphor interpretation task. They were paid for their participation. Two other groups served as judges in the scoring tasks: (1) 5 advanced undergraduate psychology students at the University of California at San Diego (U.C.S.D.), and (2) 22 undergraduate students, also from U.C.S.D. Both groups received course credit for participating.

b. Materials. The eight metaphors used were taken from Ortony (1979) and are shown in Table II. Two sets of metaphors were constructed, each containing half the metaphors in forward order (e.g., "sermons are like sleeping pills") and half in reverse order (e.g., "sleeping pills are like sermons"). Each set also contained eight filler metaphors that were always in forward order, i.e., in the most intuitively natural order. Half of the subjects saw each set so that the order assignment was counterbalanced.

[7]We stress that these features can include relations as well as object attributes. Ortony specifically mentions schemas as an instance of the kinds of representations he means the theory to apply to. However, although salience imbalance allows different kinds of predicates, distinctions among predicate types do not enter into the theory.

[8]Note that the scopes of the two theories are somewhat different. Structure-mapping makes strong predictions about aptness, not not about metaphoricity. Salience imbalance makes predictions about metaphoricity but not about aptness.

TABLE II

MATERIALS USED IN EXPERIMENT 1

Blood vessels are like aqueducts.
Surgeons are like butchers.
Education is like a stairway.
Sermons are like sleeping pills.
Cigarettes are like time bombs.
Science is like a glacier.
Encyclopedias are gold mines.
Billboards are like warts.

The eight experimental and eight filler metaphors yielded 32 object terms for the object-description task.

c. Procedure. Subjects were first asked to write descriptions of each of the individual terms (e.g., sermons, sleeping pills). The 32 object terms were presented in random order, except that the 2 terms from a metaphor were never presented contiguously. Subjects were not told about the metaphor task until they had completed the object descriptions. Then they were given the 16 metaphors in random order, in workbooks, one to a page. They were told to write the intended meaning of each metaphor and to rate its metaphoricity and aptness on separate 1–5 scales. "Metaphoricity" concerned whether the comparison was literal or nonliteral, and "aptness" concerned how clever, interesting, and worthwhile the comparison was.

d. Scoring. To test the structure-mapping hypothesis, the relationality and attributionality of the metaphor interpretations and object descriptions were rated in two ways: (1) by 5 trained advanced undergraduates *(judges' ratings)* and (2) by a group of 22 undergraduate subjects with no special training *(undergraduate A/R ratings)*. To test the salience imbalance hypothesis, 2 of the trained judges rated whether the propositions that occurred in the metaphor interpretations occurred early or late (if at all) in the object descriptions *(salience ratings)*.

e. Judges' Ratings of Relationality and Attributionality. All five judges had some advanced training in linguistics or psycholinguistics. In addition, they received roughly 10 hours of training in the use of propositional notation to represent meaning. The judges were blind both to the subjects' aptness and metaphoricity ratings and to the direction of the original metaphors. Only one judge knew the hypotheses of the study.

Three to five judges participated in each scoring session. The 20 interpretations for a given metaphor (10 from the forward presentation and 10 from the reverse) were read aloud in random order. Each judge rated the relationality and attributionality of the interpretation on separate 1–5 scales. *Relationality* was defined as the degree to which the predicates in the response expressed relations, either between *objects* (e.g., *X* hits *Y*) or between *relations* (e.g., *X* hitting *Y* causes *Y* to break). *Attributionality* was defined as the degree to which the predicates described objects in and of themselves. In most cases the decisions were straightforward. However, there were some cases in which deriving the conceptual structure from the surface information required a subjective decision. Section V gives a detailed discussion of the scoring. After rating all 20 responses for a given metaphor, the raters discussed their ratings and disagreements were resolved. The agreement before discussion was .91. Immediately after rating a metaphor interpretation, the judges rated the relationality and attributionality of the object descriptions for the same metaphor (20 descriptions for each of the two objects). Descriptions were read to the judges in a different random order from the metaphors.

f. Undergraduate A/R Ratings. As a check on the judges' ratings, a second method of scoring for relationality and attributionality was also used. This method differed from the previous rating method in three ways: (1) a large group of untrained subjects served as raters, (2) the metaphor interpretations were broken into individual propositions rather than being rated as a whole, and (3) one combined rating scale was used rather than separate scales for attributionality and relationality. Only propositions from the metaphor interpretations were rated; the object descriptions were not included in this task.

The 22 raters were divided into two groups, each scoring responses for one of the two sets of metaphors. The propositions were read to the raters in random order within and across metaphors. Each proposition was rated on a 5-point composite scale ranging from 1 (highly attributional) to 5 (highly relational). Raters were given as examples of clearly attributional statements: "*X* is red" and "*X* is large." Examples of relational statements were "*X* hits things" and "*X* causes explosions."

g. Scoring for Salience Imbalance. Two trained judges rated the metaphor interpretations for salience imbalance. They were unaware of the hypothesis being tested and of the original subjects' aptness and metaphoricity ratings. Forward and reversed metaphors were scored separately. Judges compared each metaphor interpretation with the subjects' object descriptions and assessed whether the metaphor interpretation contained any propositions also found in either the base or target object descriptions. Judges scored on the basis of meaning, not for identical wording. The outcome of this scoring procedure was, for each metaphor, the number of propositions that the original subjects had included both

in the metaphor interpretation and in (1) the base, (2) the target, (3) the top half of the base, (4) the bottom half of the base, (5) the top half of the target, and (6) the bottom half of the target.

2. *Results*

a. Structure-Mapping. The first two predictions are (1) that the metaphor interpretations should contain more relational information than attributional information and (2) that this relational advantage should be greater for the metaphor interpretations than for the object descriptions. Table III shows a typical response. Both relations and object attributes appear in the object descriptions, but only relational information appears in the metaphor interpretation.

TABLE III

SAMPLE RESPONSE IN EXPERIMENT 1: OBJECT DESCRIPTIONS AND METAPHOR INTERPRETATION OF "CIGARETTES ARE LIKE TIME BOMBS"

	Response	Trained judges' ratings: Relationality	Trained judges' ratings: Attributionality
Base:	time bomb Explosive devices with detonator linked to timing device Explosion time can be pre-set Perpetrator doesn't have to be present	5	5
Target:	cigarette Chopped cured tobacco in paper roll With or without a filter at the end held in the mouth With or without menthol Lit with a match and breathed through to draw smoke into the lungs Found widely among humans Known by some cultures to be damaging to the lungs Once considered beneficial to health	5	5
Metaphor:	cigarettes are like time bombs They do their damage after some period of time during which no damage may be evident	5	1
Aptness:	3		
Metaphoricity:	5		

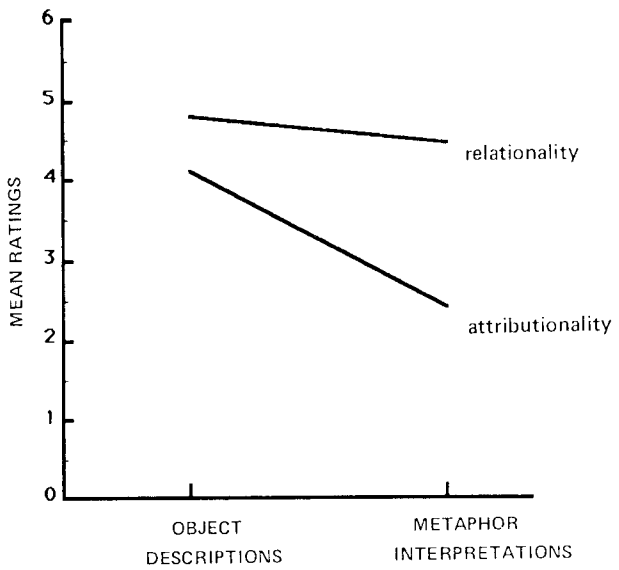

Fig. 2. Experiment 1: mean ratings by trained judges of relationality and attributionality of object descriptions and metaphor interpretations.

These predictions are borne out by the trained judges' ratings of metaphor interpretations and object descriptions. First, as shown in Fig. 2, the metaphor interpretations were rated as highly relational but not highly attributional [$t(15) = 6.68$, $p < .0005$, one tailed]. This difference holds up for individual metaphors. The mean relationality rating was higher than the mean attributionality rating for every one of the 16 metaphors, both forward and reverse.

Second, this difference between relationality and attributionality was specific to metaphor interpretations. The object descriptions were rated as high in both relational and attributional information. A 2 × 2 × 2 analysis of variance for the within-subjects factors of Directionality (forward vs. reverse), Task (metaphor vs. object), and Measure (relationality vs. attributionality) confirmed this prediction. There was a main effect of Task, simply reflecting that more was said (both attributes and relations) about the objects than the metaphors [$F(1,19) = 262.44$, $p < .001$]. Measure was also significant [$F(1,19) = 419.08$, $p < .001$], reflecting that, overall, the responses were judged as higher in relationality than in attributionality. There was no main effect of Direction [$F(1,19) = 3.20$, not significant], although a significant interaction between Direction and Task was found [$F(1,19) = 11.30$, $p < .01$]. Not surprisingly, Direction affected metaphors but not objects.

Turning to the chief result, the predicted interaction of Task and Measure was significant [$F(1,19) = 129.94$, $p < .001$]. It appears that the drop in

attributionality from object descriptions to metaphors is steeper than the drop in relationality. However, planned comparisons revealed that both attributionality and relationality differed significantly between metaphors and objects [t (39) = 18.01, $p < .001$ and t (39) = 2.05, $p < .05$, respectively]. An item analysis revealed the same patterns of significance as the subjects analysis, except that the interaction between Direction and Task was not significant. Again, the key interaction of Task and Measure was significant [F (1,7) = 15.10, $p < .01$].

The third prediction of structure-mapping is that the aptness of the metaphors should be positively correlated with the relationality of the metaphor interpretations. In contrast, there should be no correlation, or even a negative correlation, between aptness and attributionality. This prediction was confirmed using both the judges' ratings and the undergraduate A/R ratings. Figure 3 shows a scatter plot of the subjects' aptness ratings plotted against the judges' ratings of relationality and attributionality. Figure 4 shows aptness plotted against the undergraduate A/R ratings.

Pearson's product-moment correlations were performed on the mean ratings for the 16 metaphors. As predicted, aptness is positively correlated with both measures of relationality ($r = .65$, $p < .01$ for the judges ratings and $r = .56$, $p < .05$ for the undergraduate A/R rating). There is no positive correlation between aptness and attributionality, and the trend is negative ($r = -.31$, not significant) for the judges' ratings. These results suggest that subjects consider metaphors apt when they find relational interpretations.

Finally, as a check on the reliability of the measures, correlations were performed between the judges' mean ratings and the undergraduate A/R ratings. The measures proved to be consistent. The correlation with A/R rating was positive for relationality and negative for attributionality [$r(14) = .62$, $p < .05$ and $r(14) = -.65$, $P < .01$, respectively].

b. Salience Imbalance. The first prediction of salience imbalance is that the metaphor interpretations should primarily include propositions that are of high salience (mentioned early) in the base description and of low salience (mentioned late) in the target. In order to give the hypothesis every possible opportunity, several variants of the predictions were tested. Table IV shows the results. The most straightforward prediction is that more assertions from the metaphor interpretations should be found in the top half of the descriptions of the base object and the bottom half of the descriptions of the target object than from the reverse intersection (the bottom half of the base and top half of the target). Figure 5 shows a schematic depiction of this prediction ("B1" refers to top half of base and "T2" refers to bottom half of target). This prediction is not confirmed, as shown in Table IV. The difference in the mean numbers of propositions in the two intersections is not significant [$t(15) = .81$, not significant].

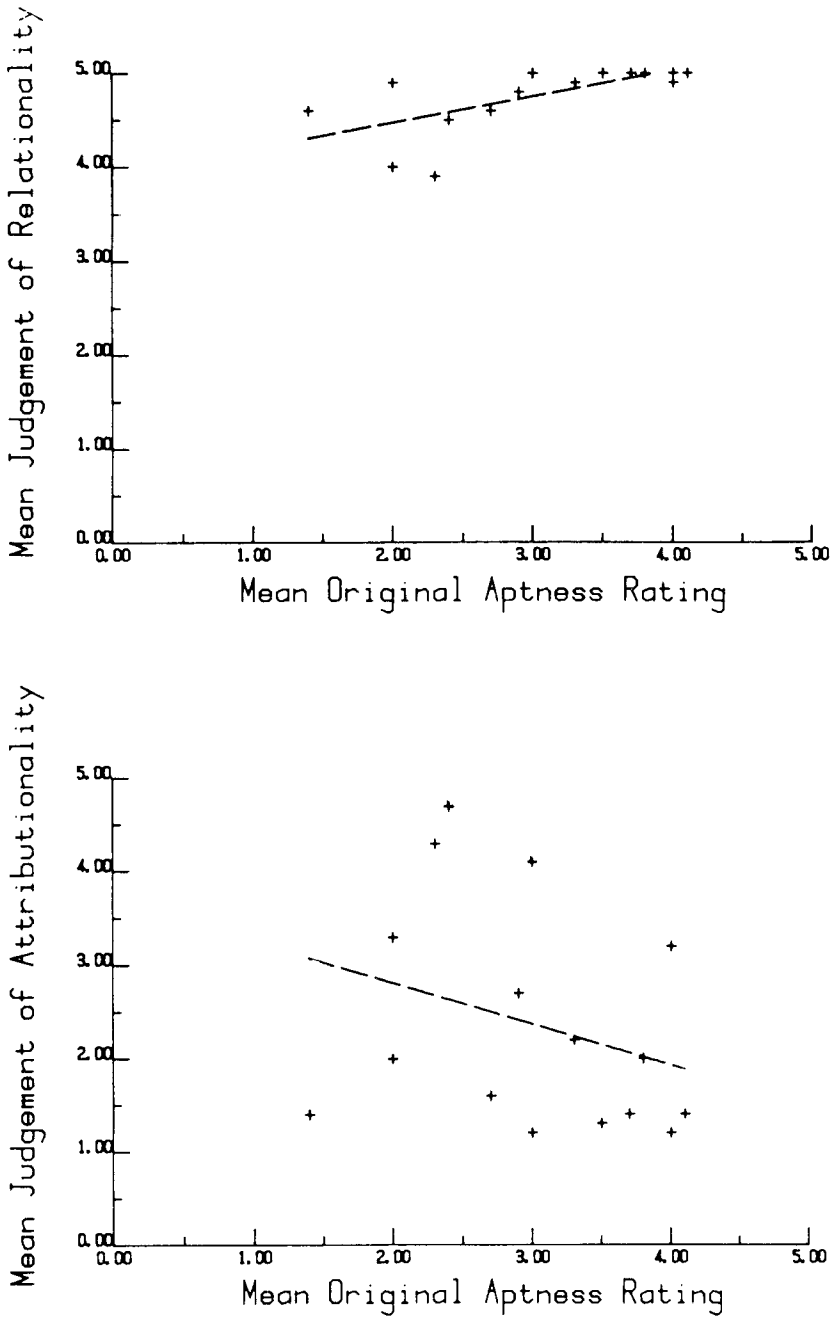

Fig. 3. Experiment 1: correlations between subjects' aptness ratings and trained judges ratings of relationality and attributionality. Top, relationality vs. aptness; bottom, attributionality vs. aptness.

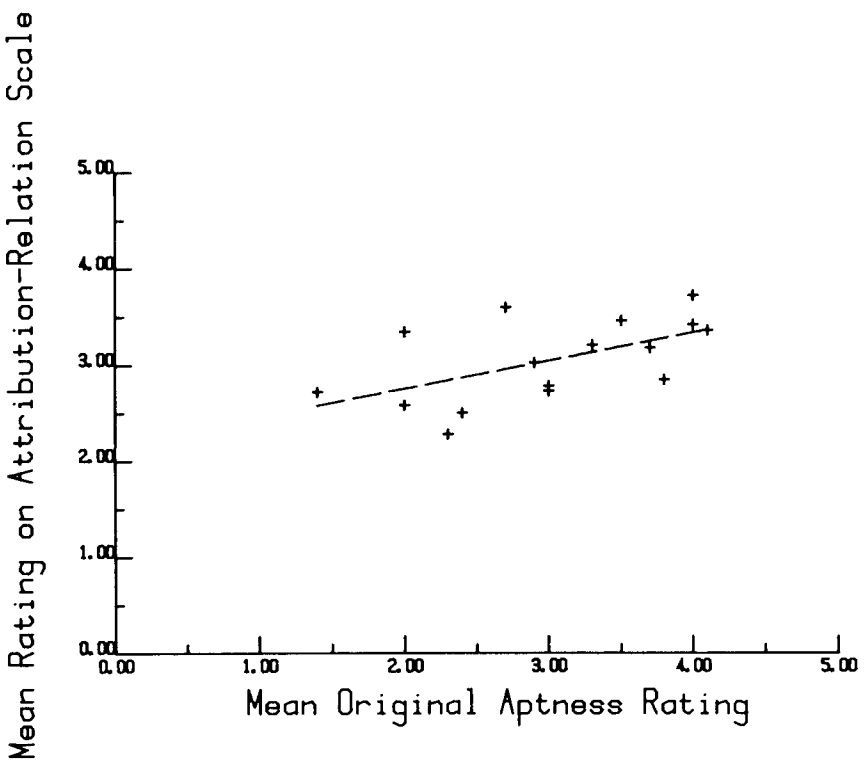

Fig. 4. Experiment 1: correlation between subjects' aptness ratings and the undergraduate A–R ratings.

TABLE IV

RESULTS OF EXPERIMENT 1: MEAN NUMBERS OF PREDICATES OCCURRING IN METAPHOR INTERPRETATIONS

Predictions of salience imbalance	Results: mean number of predicates[a]		
B1 ∩ T2 > B2 ∩ T1	B1 ∩ T2 = .038	B2 ∩ T1 = .025	NS
B > T	B = 1.16	T = 1.04	NS
B1 > B2	B1 = .58	B2 = .58	NS
T2 > T1	T2 = .49	T1 = .56	NS
B1 > T1	B1 = .58	T1 = .56	NS
T2 > B2	T2 = .49	B2 = .58	NS

[a]NS, Not significant, two-tailed t test.

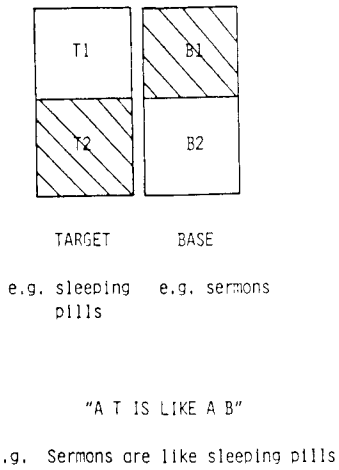

Fig. 5. Experiment 1: schematic depiction of the salience imbalance prediction: Metaphor interpretations should include information from the hatched quadrants. (T2 = bottom half of target object descriptions; B1 = top half of base object descriptions).

Perhaps the halfway point is the wrong cutoff for high vs. low salience. All or most of the information subjects mentioned in their object descriptions may be of high salience. In that case the prediction is simply that more of the metaphor assertions should match assertions from the base description than from the target description. This prediction is also disconfirmed [$t(15) = .51$, not significant]. Although two plausible versions of the salience imbalance prediction have been disconfirmed, there remain four other patterns that could support salience imbalance: if the assertions from the metaphor interpretations (1) match more assertions from the top half of the base than from the bottom half of the base; (2) match more assertions from the bottom half of the target description than from the top half; (3) match more assertions in the top half of the base than from the top half of the target; or (4) match more assertions from the bottom half of the target than from the bottom half of the base. Each one of these predictions is disconfirmed. As shown in Table IV, the relevant means are nearly identical in all cases and all differences are not significant. Overall, the first prediction of the salience imbalance hypothesis is not supported here. We found no evidence that salience imbalance determines the information people use in their metaphor interpretations.

The second prediction of the salience imbalance hypothesis, that metaphoricity ratings should be higher for forward metaphors than for reversed metaphors, also failed to receive support. Table V shows the mean aptness and metaphoricity ratings (as well as the ratings of relationality and attributionality) for forward vs. reversed metaphors. As Table V suggests, the mean metaphoricity ratings are not significantly different for forward and reversed metaphors [$t(7) = 1.21$, not significant]. The forward

TABLE V

RESULTS OF EXPERIMENT 1: COMPARISON OF FORWARD AND REVERSED METAPHORS

	Original ratings		Characteristics of interpretations		
	Aptness	Metaphoricity	Relationality (trained judges)[a]	Attributionality (trained judges)	A/R rating (group raters)
Forward metaphors	3.31	3.80	4.91	2.51	3.10
Reversed metaphors	2.70	3.60	4.60	2.24	2.99

[a]The only significant difference between the forward and reversed conditions is in relationality [$t(7) = 2.51$, $p > .05$, one tailed].

and reversed metaphors do appear to differ more in aptness than in metaphoricity, although the aptness difference is also not significant [$t(7) = 1.77$]. The only significant difference between forward and reversed metaphors is in relationality as related by the trained judges [$t(7) = 2.51$, $p < .05$]. This difference in relationality is evidence for some directional asymmetry. However, there is no evidence that this asymmetry involves differences in metaphoricity.

The third prediction of salience imbalance is that metaphoricity should be correlated with the degree of salience imbalance in the metaphors. That is, metaphoricity should be positively correlated with the proportion of metaphor interpretation statements also found in the base descriptions and negatively correlated with the proportion of interpretation statements found in the target descriptions. Instead, we find that metaphoricity is negatively correlated both with the number of statements from the target [$r(14) = -.69$, $p < .01$] and with the number of propositions from the base [$r(14) = -.56$, $p < .05$].

Since this is a key prediction for the salience imbalance theory, it seemed advisable to check if it held for the forward metaphors only. However, here, too, the results fail to show a positive correlation between metaphoricity and number of statements from the base and, indeed, show a negative trend [$r(6) = -.65$, not significant]. Finally, we tested whether the salience imbalance intuition might apply to aptness rather than to metaphoricity. This possibility also was disconfirmed. The correlations between aptness and number of propositions found in either the base or the target object descriptions are nonsignificant [$r(14) = .28$ and $r(14) = .05$, respectively].

3. *Discussion*

The results provide no support for the proposals that (1) salience imbalance determines the interpretation of a metaphor from its terms or that (2) salience imbalance is a principal source of metaphoricity. Contrary to the first proposal, the imbalance in salience of assertions in the base and target object descriptions did not predict metaphor interpretations. Contrary to the second proposal, no difference in metaphoricity was found between the forward and reversed metaphors.[9] More importantly, metaphoricity was not correlated with the degree of salience imbalance.

The predictions of structure-mapping theory were confirmed. First, the metaphor interpretations were rated higher in relationality than in attributionality. Second, this relational advantage applied specifically to the metaphors; the object descriptions were rated high in both relational and attributional information. Third, aptness was correlated with relationality. Thus, the more relational information people can find to map from the base to the target, the more apt they find the metaphor. These results are evidence that relational structure serves as a selection constraint: when people interpret metaphorical comparisons, they tacitly assume that relational information, rather than information about object attributes, is meant to be preserved in the metaphor interpretation.

One limitation of the research so far is that only eight metaphors (and their eight reversed counterparts) have been used. The small size of the stimulus set makes us suspect the null results for salience imbalance. It seemed advisable to replicate the study with a larger set of stimuli.

B. EXPERIMENT 2

Experiment 2 provides a second test of the predictions of structure-mapping theory and salience imbalance theory.[10] The basic method was the same as in Experiment 1: subjects first gave object descriptions and

[9]It should be noted that the test of this prediction is problematic. In an effort to ensure fairness to the salience imbalance position, the metaphors were taken from the set of examples that Ortony (1979) had used to illustrate the theory. However, A. Ortony (personal communication, 1986) states that at least one of these metaphors is reversible, in which case the predicted directional asymmetry in metaphoricity would not be expected to hold for that metaphor. Thus, the failure to find an overall directional asymmetry is not conclusive. However, the other two tests of salience imbalance and, in particular, the negative results concerning salience imbalance as an interpretation rule are not affected.

[10]This experiment was conducted as part of a larger study of the development of analogy and metaphor, and for this reason the metaphors were designed to be intelligible to children. For the present purposes only the adult data are of interest. See Gentner (1988b) for a presentation of the developmental findings.

then interpreted metaphors and rated them for aptness and metaphoricity. The interpretations were then scored by independent judges for relationality, attributionality, and salience, as estimated by order of mention with respect to the base and target. A key aspect of this experiment was that new materials were used which represented three types of metaphors, as discussed earlier: attributional metaphors, relational metaphors, and double metaphors. In attribute metaphors, the predicates shared by the base and target objects were object attributes: e.g., "pancakes are nickels" (both are round). In relation metaphors, the shared predicates were relations: e.g., "a tire is a shoe" (both are used by moving figures as points of contact with the ground). In double metaphors, both attributes and relations were shared: e.g., "plant stems are like drinking straws" (both are long and cylindrical; both are used to bring liquids from below to nourish a living thing).

In addition to broadening the range of materials, this collection of metaphor types provides some new tests of structure-mapping. First, since the theory predicts a correlation between aptness and relationality, the attribute metaphors should be judged as less apt than the relational metaphors. More important, for our purposes, is the test of the interpretation rules provided by the double metaphors. The results of Experiment 1 indicated that the metaphor interpretations were predominantly relational. However, it is possible that these results were obtained because only relational commonalties were possible. Since the double metaphors are designed to allow either a relational or an attributional interpretation, they provide a stronger test of the structure-mapping prediction that metaphor interpretations should be based on relations.

Structure-mapping makes three predictions. First, the metaphor interpretations should be higher in relationality than in attributionality. (This prediction applies only to the relational and double metaphors since the attribute metaphors do not permit a relational interpretation.) Second, the aptness ratings should be positively correlated with the relationality of the metaphor interpretations. Third, aptness should be lower for attribute metaphors than for relational and double metaphors.[11]

The predictions of salience imbalance are as in Experiment 1, except that in Experiment 2 the metaphors were presented in only one direction (the "forward direction"). Therefore, the predictions concerning metaphoricity and direction do not apply here. The first prediction is that metaphor interpretations should be determined by salience imbalance: that is, they should tend to include propositions mentioned early in the description of the base object and late (if at all) in the description of the target object. Second, metaphoricity should depend principally on salience imbalance: that is, the metaphoricity ratings should be positively correlated with the degree of salience imbalance.

[11] In this study the object descriptions were not rated for relationality and attributionality, so no comparisons were made between the relationality of the object descriptions and the metaphor interpretations.

1. *Method*

a. Subjects. The subjects were 10 college students from psychology classes at the University of California at San Diego.

b. Materials. As shown in Table VI, there were eight instances each of three metaphor types: (1) attribute metaphors, in which base and target shared many attributes but few relations; (2) relation metaphors, in which base and target shared many relations but few attributes; and (3) double metaphors, in which base and target shared both relations and attributes. All subjects interpreted all 24 metaphors. In addition, there were 48 objects in the object description task.

TABLE VI

MATERIALS USED IN EXPERIMENT 2

Relational metaphors	The moon is like a lightbulb.
	A camera is like a tape-recorder.
	A ladder is like a hill.
	A cloud is like a sponge.
	A roof is like a hat.
	Treebark is like skin.
	A tire is like a shoe.
	A window is like an eye.
Attributive metaphors	Jellybeans are like balloons.
	A cloud is like a marshmallow.
	A football is like an egg.
	The sun is like an orange.
	A snake is like a hose.
	Soap suds are like whipped cream.
	Pancakes are like nickels.
	A tiger is like a zebra.
Double metaphors	A doctor is like a repairman.
	A kite is like a bird.
	The sky is like the ocean.
	A hummingbird is like a helicopter.
	Plant stems are like drinking straws.
	A lake is like a mirror.
	Grass is like hair.
	Stars are like diamonds.

c. Procedure. The procedure was the same as for Experiment 1. Subjects first wrote out descriptions of the 48 separate objects, which were presented in random order. They were then given the metaphor workbook and told to write their interpretations of the metaphors and to rate their aptness and metaphoricity.

d. Scoring. The metaphor interpretations were scored by trained judges (the same five advanced undergraduates as in Experiment 1). The method was as in Experiment 1: groups of from two to four judges were read the metaphor interpretations and rated them on two 5-point scales, a relational scale and an attributional scale. Agreement between raters ranged from 85 to 100% on different metaphors.

2. *Results*

The results of this study largely replicated those of Experiment 1.

a. Structure-Mapping. The first prediction is that the interpretations of relation metaphors and double metaphors would be high in relational information but not attributional information. As discussed previously, the performance on double metaphors is of special interest since they were specifically designed to have attributional interpretations as well as relational interpretations. Table VII shows the rated relationality and attributionality of the interpretations for the three types of metaphor. As predicted, metaphor interpretations were rated higher in relationality than in attributionality for both relation and double metaphors [$t(9) = 5.87$, $p < .001$ and $t(9) = 3.79$, $p < .05$, respectively].

The second prediction of structure-mapping is that aptness ratings would be positively correlated with relationality, but not with attributionality.

TABLE VII

RESULTS OF EXPERIMENT 2: JUDGES' RATINGS OF THE RELATIONALITY AND ATTRIBUTIONALITY OF THE METAPHOR INTERPRETATIONS AND SUBJECTS' APTNESS RATINGS FOR THE THREE TYPES OF METAPHORS

	Relationality	Attributionality	Aptness
Relational metaphors	4.6	1.7	2.9
Double metaphors	3.9	3.0	2.9
Attributional metaphors	1.4	4.4	2.3

As in Experiment 1, aptness is positively correlated with relationality [$r(22) = .55$, $p < .01$]. However, the attributionality results are somewhat stronger than those for Experiment 1 in that aptness is *negatively* correlated with attributionality [$r(22) = -.42$, $p < .05$]. It appears that subjects found metaphors more apt to the extent that they found relational commonalities and less apt to the extent that they found attributional commonalties.

The third prediction is that subjects should consider the relation and double metaphors more apt than the attribute metaphors since the first two permit relational interpretations and the latter does not. As Table VII shows, this prediction is confirmed; the mean aptness ratings for relation and double metaphors are considerably higher than those for attribute metaphors [$t(9) = 5.24$, $p < .001$ and $t(9) = 7.31$, $p < .001$, respectively]. Thus, all three predictions of structure-mapping are confirmed.

b. Metaphor Classes. One other set of findings concerns the materials. Crucial to this theory is the claim that the distinction between attributionality and relationality can be made reasonably clearly. The results provide evidence for the orderliness of the distinction (see Table VII). First, the relation metaphors were rated as more relational than both attribute and double metaphors [$t(9) = 12.98$, $p < .001$ and $t(9) = 2.35$, $p < .05$, respectively]. Second, the attribute metaphors were rated as more attributional than both relation and double metaphors [$t(9) = 18.09$, $p < .001$, and $t(9) = 8.11$, $p < .001$, respectively]. Finally, the double metaphors were intermediate on both rating scales. They were more relational than attribute metaphors [$t(9) = 13.12$, $p < .001$] and more attributional than relation metaphors [$t(9) = 10.87$, $p < .001$].

c. Salience Imbalance. The first prediction of salience imbalance is that the metaphor interpretations should tend to include propositions mentioned early in the description of the base and late in the description of the target. This result is not confirmed. Table VIII shows the results for the same set of six versions of the prediction tested in Experiment 1. Not one yields a significant difference. Indeed, the pattern of results is remarkably similar to the negative results of Experiment 1. Thus, it does not appear that salience imbalance functioned as an informational constraint on which propositions subjects included in their metaphor interpretations.

The second prediction of salience imbalance is that metaphoricity should depend positively on salience imbalance. That is, metaphoricity should be positively correlated with the number of assertions from the metaphor interpretation that match with the base description and negatively correlated with the number that match with the target. This prediction, too, is not confirmed. The correlations between rated metaphoricity and number of propositions from base and from target are not significant [$r(22) = .32$ and $r(22) = .10$, respectively].

TABLE VIII

RESULTS OF EXPERIMENT 2: MEAN NUMBERS OF PREDICATES OCCURRING IN METAPHOR INTERPRETATIONS

Predictions of salience imbalance	Results: mean number of predicates[a]		
B1 ∩ T2 > B2 ∩ T1	B1 ∩ T2 = .025	B2 ∩ T1 = .026	NS
B > T	B = .63	T = .59	NS
B1 > B2	B1 = .37	B2 = .28	NS
T2 > T1	T2 = .23	T1 = .39	NS
B1 > T1	B1 = .37	T1 = .39	NS
T2 > B2	T2 = .23	B2 = .28	NS

[a]NS, Not significant, two-tailed *t* test.

As in Experiment 1, we also tested the correlation between salience imbalance and aptness, even though that correlation is not predicted by the salience imbalance model. In Experiment 2, unlike Experiment 1, there was some support for a correlation between salience imbalance and aptness. That is, aptness was positively correlated with the number of propositions from the base [$r(22) = .37$, $p < .01$]. However, the further prediction that aptness should be negatively correlated with the number of propositions from the target was not confirmed [$r(22) = .30$, not significant].

3. *Discussion*

The results are consistent with the structure-mapping claim that relational selectivity acts as an informational constraint on the interpretation of a metaphor from its terms. First, the metaphor interpretations were rated high in relationality and low in attributionality, including the double metaphors, which could support either a relational or an attributional interpretation. Second, the aptness ratings were positively correlated with judged relationality and negatively correlated with judged attributionality. Finally, subjects rated the relational and double metaphors as more apt than the attribute metaphors. Subjects appear both to seek relational predicates in metaphor interpretation and to judge the aptness of the comparison according to the relationality of the interpretation.

No support was found for the two predictions of salience imbalance tested here. First, the relative salience of assertions in the base and target object descriptions did not predict inclusion in the metaphor interpretations. Second, there was no correlation between degree of salience imbalance and metaphoricity. (However, some evidence was obtained for a relation between salience imbalance and aptness. We will return to this point later.)

Both structure-mapping and salience imbalance postulate dependent variables that could be operationalized in various ways. In Experiment 3a and 3b we consider alternate methods of assessing the predictions of the two theories. In Experiment 3a we consider an alternative way to operationalize the structural distinctions necessary to test structure-mapping theory. In the previous experiments, we used two methods of rating the underlying propositional structure of subjects' responses. In both cases the results confirmed the structure-mapping predictions. However, these methods have the disadvantage that decisions about attributionality and relationality are, in part, subjective. Although the judges were not told the subjects' aptness or metaphoricity ratings, and among the five trained judges only one knew the hypothesis, it nevertheless seems desirable to have a less subjective measure of relationality and attributionality. Therefore, in Experiment 3a the structure-mapping predictions were evaluated using a syntatic scoring method. The task of the raters was simply to assign each word to a syntactic category. Then the categories were sorted into relational categories (e.g. transitive verbs and comparative adjectives) and object-attribute categories (e.g., common nouns and adjectives). Our reasoning was that the number of relational words in the surface syntax might be fairly well correlated with the relationality of the underlying propositional structure. If so, this syntactic scoring system should produce results similar to the two propositional scoring systems. Although this method entails some sacrifice in the richness of propositional scoring (that is, it will fail to capture all of the underlying relational structure), it has the advantage of being objective and easily describable.

C. EXPERIMENT 3A

1. *Method*

a. Materials. The descriptions of the 16 objects and interpretations of the 8 forward and reverse metaphors from Experiment 1 were given to the raters.

b. Raters. Five advanced undergraduates from the University of Illinois, who each received a brief training session in grammatical distinctions, served as raters.

c. Rating System. The raters were instructed to rate each word of every sentence used in the descriptions and interpretations according to the grammatical categories shown in Table IX. Section VI gives a detailed description of the scoring system used. Our interest was in scoring what

TABLE IX

EXPERIMENT 3A: GRAMMATICAL CATEGORIES USED IN SYNTACTIC SCORING SYSTEM

Category	Example
Noun	
Ordinary noun[a]	Dog
Relational noun	Father
Not sure (noun)	—
Pronoun	They
Adjective	
Ordinary adjective[a]	Red
Relational adjective	Edible
Not sure (adjective)	—
Pronominal adjective or possessive[b]	Their, theirs
Comparative modifier[b]	Bigger than
Preposition[b]	Above
Adverb	Slowly
Verb	
Intransitive verb	Fall
Transitive verb[b]	Hit
Linking verb	Is, seems
Verb plus particle	Hang on
Not sure (verb)	—
Connective[b]	Because
Other	—

[a]Counted as attributional in subsequent analyses.
[b]Counted as relational in subsequent analyses.

subjects wrote about the objects (or, in the case of metaphor interpretation, about the comparison between the objects). Therefore, any mention of the original object terms as well as any pronouns that referred to them was omitted from the scoring. For example, if a subject interpreted the metaphor "an apple is like a fire engine" by saying "they both are red" or "apples are red," then only the descriptor *red* would be scored; the words *apples*, *they*, and *both* would be omitted from scoring (since they merely refer to the objects given in the metaphor). Articles such as *the* and *an* were not scored.

d. Procedure. The raters worked in groups of three. After practice with filler materials, they scored the first three responses for each metaphor and object. After these ratings were recorded, disagreements were discussed. Finally, the rest of the statements were scored individually and the ratings were recorded. Any remaining disagreements were then discussed and resolved.

e. Analysis. The raters categorizations were combined into three larger scores: relational, attributional, and other. The categories counted as relational were (1) pronominal adjectives and possessives, (2) comparative modifiers, (3) prepositions, (4) transitive verbs, (5) connectives, (6) relational nouns, (7) relational adjectives, and (8) verb–particle combinations. The categories counted as attributional were ordinary nouns and ordinary adjectives. A narrow criterion was also used, in which categories 6, 7, and 8 were omitted from the relational category. Certain categories were omitted from further analysis because they are neither clearly relational nor clerly attributional: (1) not sure (nouns), (2) not sure (adjectives), (3) adverbs, (4) linking verbs (e.g., copulas), and (5) other. For each response (that is, each metaphor interpretation and each object description) the number of words falling into each class—*relational*, *attributional*, and *other*—was computed.

2. *Results and Discussion*

The results again confirm the predictions of structure-mapping. As in Experiment 1, the proportion of attributional terms used in the object descriptions was significantly greater than the proportion used in the metaphor interpretations (.38 vs. .31). In contrast, the proportion of relational terms was similar: .39 in the object descriptions and .37 in the metaphor interpretations. Thus, the proportion of attributional information dropped as subjects moved from object descriptions to metaphor interpretations. This pattern held for both forward and reverse metaphors.

A 2 × 2 × 2 analysis of variance of Direction (forward vs. reversed), Task (metaphor vs. object), and Category (relational vs. attributional) confirmed these patterns. There was a main effect of Task, reflecting the greater overall scores for object descriptions than for metaphors [$F(1,19) = 16.85$, $p < .0001$], and a main effect of Category, reflecting the overall greater amount of attribute information [$F(1,19) = 8.89$, $p < .01$]. Most importantly, there was an interaction between Task and Category, reflecting the fact that the proportion of attributional, but not relational, information was greater in the object descriptions than in the metaphor interpretations [$F(1,19) = 5.56$, $p < .05$]. No other effects were significant, including that of Direction.[12] Thus, the syntactic scoring system yields the same pattern of results, supporting the structure-mapping view, as was originally obtained by the propositional scoring system in Experiment 1.

Having tested the predictions of structure-mapping under a different

[12]The results of the narrow summation were similar, although much compressed, since by this scheme about half the terms had to be omitted from the analysis. The pattern of significance was the same except that the main effect of Category was nonsignificant.

method of assessment, in Experiment 3b we similarly reexamine the predictions of salience imbalance. The results of Experiments 1 and 2 failed to provide support for salience imbalance. However, before accepting these negative results, we must consider the possibility that the order-of-mention method used here simply failed to provide an adequate indicator of salience. Order of mention may be affected by multiple variables and, therefore, may be insensitive to the true salience order. It should be noted that some aspects of the data argue against this interpretation. First, the detailed patterns of negative results with respect to metaphoricity are nearly identical for Experiments 1 and 2. Second, the salience imbalance predictions fail not only on the detailed comparisons (e.g., top half of base vs. bottom half of base) but also on the simple comparison of the relative contribution of base vs. target. By any reasonable interpretation of the notion of salience, it would seem that subjects should have included at least some information in their object descriptions that they considered salient for the objects. Yet in neither experiment did the base contribute more to the metaphor interpretation than the target. Thus, the salience imbalance predictions fail both at the fine-structure level and at the global level of base versus target. However, despite these arguments, it remains the case that order-of-mention is only one measure of salience. Therefore, in Experiment 3b we used an alternative way of measuring salience.

D. Experiment 3b

In Experiment 3b, metaphor interpretations from Experiment 1 were assessed for salience relative to the base and target terms using a method similar to that used by Ortony *et al.* (1985). Subjects rated the assertions in the metaphor interpretations for their *immediacy* (how readily the thought comes to mind when thinking about the object term) and *importance* (how important or significant the thought is with respect to the object term) with respect to the base and target terms. Salience imbalance predicts that the assertions from the metaphor interpretation will be rated as more immediate and important with respect to the base than with respect to the target.

Our interest is in whether metaphor interpretations can be predicted from the salience of information in the *prior* representations of base and target objects. Therefore, our method differs from that of Ortony *et al.* in that subjects were not told about the metaphor that generated the assertions they rated. This was done to ensure that subjects would give the immediacy and importance of the assertions simply with respect to the relevant object terms unbiased by the context of the metaphor. A second difference in methodology was that subjects rated assertions from the metaphor interpretations in the context of other assertions from the object descriptions. This meant that salience of the metaphor assertions was assessed relative to the salience of other assertions about the base and target terms.

1. Method

a. Subjects. The subjects were 77 undergraduate and graduate students of the University of Illinois. They either received class credit or were paid for their participation.

b. Materials. Subjects rated statements both from the metaphor interpretations (both forward and reverse metaphors) and from the object descriptions given by subjects in Experiment 1. Each subject rated the immediacy and importance of three types of statements with respect to one of the terms of the original metaphor: (1) statements about the term itself, (2) statements from metaphors in which the term was the target, and (3) statements from metaphors in which the term was the base. For example, for the metaphor "blood vessels are like aqueducts," some subjects rated statements with respect to "blood vessels" and others rated statements with respect to "aqueducts." Subjects in the "blood vessels" group rated the immediacy and importance, with respect to blood vessels, of (1) all statements describing blood vessels, (2) all statements made in interpreting the forward metaphor "blood vessels are aqueducts," and (3) all statements made in interpreting the reverse metaphor "aqueducts are blood vessels."[13]

The statements were rated with respect to the 16 terms that entered into the metaphors in Experiment 1. The number of statements about a given term varied from 46 to 73 depending on how many statements were made by the original subjects. Statements that were repeated across subjects were rated only once. Repetitions were scored on the basis of similar or identical meaning. For example "they transport fluid" as a description of "aqueducts" was counted as a repetition of "they carry liquid." The weighted scoring system described next took into account the frequency of the original statements. To keep subjects unaware of the metaphors, we removed any mention of the original terms by Experiment 1 subjects. For example, if the metaphor had been interpreted as "both blood vessels and aqueducts provide nutrients," this statement was changed to "they provide nutrients." This had the further advantage that it allowed the identical statements to be rated with respect to both "blood vessels" and "aqueducts."

c. Procedure. Subjects were told that they would hear statements about a particular topic. Their job was to rate each statement on scales

[13] A few statements that were clearly inappropriate were omitted. For example, one subject described "aqueduct" as "a horse racing track in Queens, NY." This response was omitted.

of immediacy and importance with respect to that topic. *Importance* was defined as "how important or significant the thought is with respect to the (topic)." Importance was rated on a scale from 1 (not at all important) to 5 (highly important). *Immediacy* was defined as "how quickly and naturally the thought comes to mind when you are thinking of (the topic)" and was also rated on a scale from 1 to 5. As discussed previously, statements from the object descriptions were interspersed with statements from the interpretations of the metaphors that included that object.

Each term was rated by 10 subjects. Of the 77 subjects, 56 rated two terms, 10 rated one, 8 rated three, 2 rated four, and 1 rated six.[14] Subjects never rated two terms taken from the same metaphor.

d. Scoring. For each term, the ratings of the metaphor interpretation statements were summed separately for the forward and reverse metaphors. This gave us two totals across subjects, one for assertions about metaphors in which the term functioned as the base object and one for metaphors in which the term functioned as the target object. Both weighted and unweighted totals were computed. For the unweighted totals, the raw ratings of statements across subjects were simply summed for each term. For the weighted totals, the rating for each statement was weighted by the number of times the statement had been given by the original subjects.

2. *Results and Discussion*

If salience imbalance provides an interpretation rule for metaphors, then the statements that occurred in the metaphor interpretations should be rated as more salient with respect to base terms than target terms. As shown in Tables X and XI and Fig. 6, this prediction was not confirmed for either the immediacy or the importance ratings.

Four 2 × 2 repeated-measures analyses of variance (ANOVA) were performed over immediacy and importance using both the weighted and unweighted scores. For all four ANOVA, the two factors were Position (base vs. target) and Directionality (forward vs. reverse metaphor interpretation) with items as the random variable. The key prediction of salience imbalance is a main effect of Position: the immediacy (or importance) of assertions should be higher with respect to the base than the target. This prediction was not confirmed. The effect of Position did not reach significance in any of the four analyses. For immediacy, the weighted and unweighted analyses yielded $F(1,7) = 1.24$ and $F(1,7) = 1.23$, respectively. For importance, $F(1,7) = .69$ and $F(1,7) = 1.81$ for the weighted and unweighted analyses, respectively. Thus, for both measures of salience

[14]This task was used as a filler task for other experiments. Thus, the number of terms rated by a subject varied according to the time allotted for the filler task.

TABLE X

RESULTS OF EXPERIMENT 3B: MEAN IMMEDIACY RATINGS OF INFORMATION IN THE METAPHOR INTERPRETATIONS WITH RESPECT TO THE BASE AND TARGET

	Weighted		Unweighted	
	Base	Target	Base	Target
Forward	3.18	2.59	3.04	2.55
Reverse	2.79	3.21	2.72	3.02
Combined	2.98	2.90	2.88	2.78

neither the weighted nor unweighted ratings revealed any significant asymmetry between the salience of the metaphor interpretations with respect to the base and their salience with respect to the target.

The main effect of Directionality was not significant in any of the four analyses. Statements that occurred in the interpretations of forward metaphors were not considered either more immediate or more important than statements in the interpretations of reverse metaphors.

Although no main effects were shown in any of the four ANOVAs, there was an interaction between Position and Direction for immediacy [$F(1,7) = 13.72$, $p < .01$ for the weighted analysis and $F(1,7) = 9.18$, $p < .05$ for the unweighted analysis]. Figure 6 shows the weighted immediacy results with respect to the base and target for forward and reverse metaphors. The forward metaphors show the predicted asymmetry—that is, statements from the metaphor interpretations are rated as more immediate in the base than in the target—but the reverse metaphors show the opposite pattern. (The importance ratings showed a nonsignificant crossover of the same form.) Had salience imbalance operated as a strong interpretation

TABLE XI

RESULTS OF EXPERIMENT 3B: MEAN IMPORTANCE RATINGS OF INFORMATION IN THE METAPHOR INTERPRETATIONS WITH RESPECT TO THE BASE AND TARGET

	Weighted		Unweighted	
	Base	Target	Base	Target
Forward	3.23	2.87	3.18	2.90
Reverse	2.90	3.17	2.95	3.09
Combined	3.07	3.02	3.07	2.99

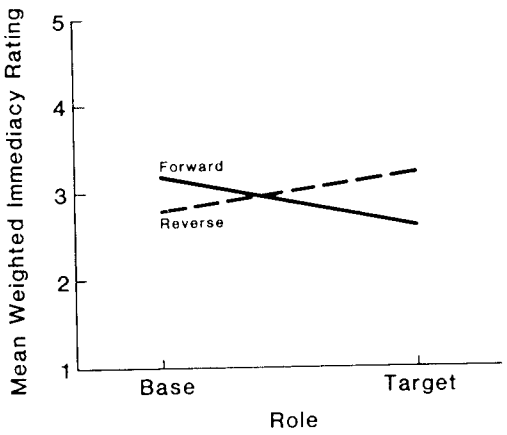

Fig. 6. Experiment 3b: weighted immediacy ratings of propositions from the interpretations of forward and reverse metaphors with respect to base and target terms.

rule, the immediacy and importance of the metaphor assertions would have been higher for the base than for the target regardless of the direction of the metaphor.

Thus, the results of this study largely confirm our previous negative findings concerning salience imbalance. Using immediacy and importance ratings as a measure of salience, we found no evidence that salience imbalance determines the information that is included in metaphor interpretations. The immediacy interaction provides support for a directional effect in that for forward metaphors the predicted asymmetry was obtained. However, the fact that the reverse metaphors show the reverse asymmetry indicates that, although salience imbalance may capture a genuine order preference, it does not determine subjects' interpretations. Indeed, the pattern of results suggests that subjects had some other informational constraint (e.g., preserving common relational structure) that determined which commonalities belonged in their metaphor interpretations and simply included these commonalities regardless of their relative salience. If we make the further assumption that the "forward" metaphors are just those in which the relational commonalities are of higher salience in the base (and therefore that the "reverse" metaphors tend to have relational commonalities that are of higher salience in the target), then the fact that forward metaphors obeyed salience imbalance and reverse metaphors did not is explained. We will return to the issue of the respective roles of structure-mapping and salience imbalance in metaphor interpretation in the following discussion.

E. General Discussion: Salience Imbalance and Structure-Mapping

1. *Structure-Mapping*

The results of the three experiments support the application of structure-mapping to metaphor. In Experiments 1, 2, and 3a, metaphor interpretations were found to contain predominantly relational information and to include relatively less attributional information than the object descriptions for their terms. Even metaphors deliberately designed to suggest either an attribute or relational interpretation (the double metaphors) received relational interpretations. This finding held under three different methods of scoring relationality and attributionality, two based on judgments of propositional structure and one based on syntactic judgments. Not only did subjects tend to focus their metaphor interpretations on relational information, but they also appeared to base their aptness judgments on the degree to which they succeed in arriving at a relational interpretation. In both Experiments 1 and 2, the relationality of metaphor interpretations was positively correlated with aptness. In contrast, subjects appeared to find attribute matches irrelevant or even detrimental to their sense of how apt a metaphor was; the correlation between aptness and attributionality was negative in Experiment 2 and not significant (but with a negative trend) in Experiment 1. Further, in Experiment 2 subjects judged the aptness of the double metaphors according to the relationality of their interpretations, even though an attribute interpretation was also clearly possible, and they judged both relational and double metaphors as more apt than attribute metaphors.

2. *Salience Imbalance*

The results for salience imbalance are both less promising and more puzzling. We first consider the negative evidence which suggests that salience imbalance provides neither an interpretation rule nor a defining criterion for metaphor. We then consider some positive findings that suggest that salience imbalance does have a role in metaphor.

a. Negative Evidence Concerning Salience Imbalance. In all three experiments our methods, unlike those of Ortony *et al.* (1985), were aimed at predicting metaphor interpretations from the salience of information in the prior representation of the constituent terms. We found no evidence that the relative salience of information in the representations of the terms determines the meaning given to a metaphor. In Experiments 1 and 2, using an order-of-mention measure of salience, we found no significant

tendency for the metaphor interpretations to contain high-salient information from subjects' prior descriptions of the base terms and/or low-salient information from descriptions of target terms. The patterns of results in the two experiments are remarkably similar even though different metaphors were used. In Experiment 3, two alternative measures of salience were used. New subjects were given a pool of propositions from previous subjects' metaphor interpretations and object descriptions and rated their immediacy and importance with respect to the base or target terms. No overall asymmetry was found. In particular, contrary to the prediction of salience imbalance, there was no tendency for the ratings of assertions from the metaphor interpretations to be rated as more salient (by either measure) with respect to the base than with respect to the target. Thus, again we found no support for salience imbalance as an interpretation rule for metaphor.

The second strong construal of salience imbalance is that it is constitutive of metaphor. No evidence was found for this claim in either Experiments 1 or 2. There was no difference in the metaphoricity of forward and reverse metaphors in Experiment 1, and more importantly, no correlation was found between metaphoricity and salience imbalance, even when only forward metaphors were considered. This undermines the claim that metaphoricity depends crucially on salience imbalance. Aside from the findings reported here, a second reason to question salience imbalance as the source of the distinction between literal similarity and metaphor is that both literal and nonliteral similarity statements show directional asymmetry and salience imbalance (Rosch, 1976, 1975; Tversky, 1977; Ortony *et al.*, 1985). Thus, salience imbalance and asymmetry are not defining characteristics of metaphor. On the other hand, the evidence of Ortony *et al.* suggests that directional asymmetries are stronger in metaphor than in literal similarity comparisons. (Note, however, that Conner & Kogan, 1980, failed to find greater asymmetry for metaphor). We return to this point shortly.

b. Positive Evidence for Salience Imbalance. The strong proposals that salience imbalance provides an informational constraint on metaphor interpretations or that it is definitive of metaphor do not appear to hold. However, there are three findings that indicate that salience imbalance plays a special role in metaphor. First, the results of Ortony *et al.* suggest that salience imbalance and directional asymmetry may be stronger for metaphors than for literal similarity statements. Second, in Experiment 2, but not in Experiment 1, we found some partial evidence that salience imbalance influences aptness judgments in that there was a positive correlation between aptness and order of mention in the base, though no negative correlation was found with order of mention in the target. Third, the results of Experiment 3 showed evidence for salience imbalance (as measured by immediacy ratings) for forward metaphors, though not for reverse metaphors. These findings suggest that salience imbalance may have some special status in metaphor.

c. Salience Imbalance as a Communicative Norm. Perhaps the best account of the role of salience imbalance comes from an observation of Ortony *et al.* (1985). They suggest that directionality and salience imbalance may arise from a conversational contract—a version of the "given-new" contract described by Clark and Haviland (1977). When a speaker uses a simile such as "*a* is like *b*," the hearer has a pragmatic understanding about what is likely to be conveyed: "In similes (and indeed in all similarity statements) the 'given' entity is the topic of the comparison and therefore is in the *a*-position. The 'new' information that is being communicated about the given entity is contained in the *b*-term. . . . Presumably, to convey the new information, a speaker selects a *b*-term for which the attributes to be communicated are highly salient" (Ortony *et al.*, 1985, p. 571). Similar points have been made by Glucksberg (1980) and by Tourangeau and Sternberg (1981). Thus, salience imbalance may be the application of a conversational cooperativeness rule to comparatives: "If *X* is to be explained by comparison with *Y*, then the explanation should be more accessible for *Y* than for *X*."

By this account, salience imbalance should be viewed not as an informational constraint on the interpretation of metaphor but as a communicative norm. The findings discussed here are, in the main, consistent with this account of salience imbalance. We found no evidence that salience imbalance provides an interpretation rule. However, if we assume that metaphors with high "forward" salience imbalance are more communicatively appropriate, this could account for the directional preferences found by Ortony *et al.* (1985). Such a directional preference could also account for the relation between aptness and salience imbalance for forward metaphors found in Experiment 2. This account does not explain why salience imbalance and asymmetry are more characteristic of metaphor than of literal similarity. Perhaps a further aspect of the conversational contract is that the more demanding it is to comprehend an utterance, the more important it is for it to obey good communicative norms. If we make this assumption, and further assume that metaphors are more difficult to comprehend than literal similarity statements, then we can explain the greater directional asymmetry of metaphors.

This line of argument suggests that the two theories are directed at different, though interrelated, aspects of metaphor comprehension. Structure-mapping describes a set of informational constraints on the kind of information that should enter a metaphor interpretation, while salience im-

balance describes a communicative norm about the way new information should be presented in a metaphor. That is, when people hear a metaphor they seek a common system of relational information, and they assume that this system will be more obvious in the base than it is in the target. Or, to put it more informally, structure-mapping tells you what to look for and salience imbalance tells you where to look first. An interesting prediction that arises from this account is that salience imbalance effects should be most pronounced in communication contexts, whereas structure-mapping effects should hold equally within and outside of communication contexts.[15]

The results of the three experiments provide considerable support for the predictions of structure-mapping. To return to the questions raised in the introduction, we conclude that, despite the computational advantages of a simple representational system, we cannot model metaphor interpretation without distinguishing among different kinds of predicates—specifically, between relations and attributes. In the next section we consider the further issue of higher-order relational structure in the interpretation of metaphor and analogy.

IV. Systems of Relations

Structure-mapping describes two interpretation rules for analogy and metaphor. The studies discussed so far provide support for the first of these claims, that people's interpretations tend to include relations common to both domains and tend to disregard common object attributes. We turn now to the second informational constraint postulated by structure-mapping, the *systematicity principle* which states that people implicitly seek to interpret an analogy or metaphor in terms of a shared set of interconnected relations constrained by higher-order relations rather than in terms of isolated predicates. That is, among the (potentially large) set of common relations that could be included in an interpretation, people select those that are part of a common *system* of relations. Thus, the systematicity principle acts as a selection filter that selects which lower-order predicates are preserved in an interpretation. It reflects a tacit preference for coherence and deductive power in analogy and metaphor.

In this section we discuss the role of higher-order relational structure in the interpretation of analogy and metaphor. We present empirical evidence for systematicity as an informational constraint on the interpretation of analogy.[16] Finally, we consider a competing account of the role of higher-order knowledge in analogical processing.

[15]Note that this account of salience imbalance is compatible with Ortony's (1979) third suggestion, that is, with a processing model in which an interpreter starts from the top of the base and works down, searching for components that match with the target. However, in light of the present results, we would add to this the structure-mapping constraint that the match be one of relational structure.

A. Evidence for Systematicity

Although researchers have just begun to test systematicity as a psychological interpretation rule, there is some evidence for the general importance of systematicity in analogical processing. For example, Gentner and Schumacher (1986) and Schumacher and Gentner (1988) have obtained evidence that systematicity can facilitate accurate analogical mapping. In their research, subjects were taught a device model and then asked to transfer their knowledge to an analogous device. Subjects were able to achieve accurate transfer in substantially fewer trials when their initial model possessed systematic structure than when it did not (even though the same procedures were taught in both cases). Other research indicates that common systematic structure influences subjects' judgments of how sound an analogy is (i.e., whether or not the analogy would yield justifiable inferences) (Gentner & Landers, 1985; Rattermann & Gentner, 1987). In these studies subjects rated the soundness of the analogical match between two brief stories. All the pairs of stories included matching lower-order relations. In some cases these lower-order matches were governed by matching higher-order relations; in other cases, the lower-order matches simply stood alone. Subjects rated the comparisons between stories as more sound when the higher-order matches were present. These sets of results implicate systematicity in analogical processing and in the evaluation of analogies. However, this research does not tell us whether systematicity acts as a selection filter that influences the way the meaning of an analogy or metaphor is derived from the meaning of the base and target.

Research by Clement and Gentner (1988) addresses this issue. We are conducting studies of the interpretation of analogy to examine whether systematicity constrains the selection of information to map between a base and target. Specifically, we ask if the selection of which lower-order relations to include in the interpretation is governed by whether or not these relations belong to a larger system of relations shared by the two domains. In these studies subjects read descriptions of two analogous domains and are asked to say which aspects of the target contribute to the analogy. The base and target are descriptions of fictional worlds. In each case, the base and target share two key lower-order relations (which we will call "key facts"). For example, one base domain describes the life-cycle of some extraterrestrial creatures. The two key facts in this domain

[16]The remaining experiments utilize analogies. If, as we have argued, relational metaphors are treated like analogies, the conclusions will apply to them as well as to analogies.

are (1) that the creatures periodically become dormant and (2) that they can only survive in one particular environmental niche. The analogous target domain concerns robots who explore planets. The target includes analogous versions of the two key facts described in the base: (1) the robots periodically shut down their primary functions, and (2) the robots are unable to function in different locations.

In both domains, the key facts are governed by higher-order systems of causal relations. The essential manipulation is whether or not these causal systems match between the base and the target. Thus, in one case the matching key facts are governed by matching causal systems (the matching system case). For example, the lower-order facts that the creatures/robots become dormant might be caused by like events in both the base and target, e.g., lack of necessary resources. In contrast, the other pair of matching key facts—e.g., that the creatures/robots have restricted areas of functioning—is governed by different causes in the base and the target (the non-matching-system case). Note that in both cases there is a perfect match between the lower-order predicates. The difference is that in one case, the two like lower-order relations belong to *like* systems of relations while in the other, the two like lower-order relations belong to different systems of relations. (Which of the two facts is identically governed is counterbalanced across subjects.)

Subjects study the base and the target and then choose which of the two lower-order facts in the target best contribute to the analogy with the base. Our hypothesis is that subjects should select the matching-system lower-order relation over the non-matching-system lower-order relation. The results of the initial studies support this hypothesis: subjects chose (study 1) or predicted (study 2) the matching-system fact more often than the non-matching-system fact. Note that the subjects' overt task was to choose between two pairs of lower-order relations that are, by themselves, equally well matched. Yet subjects showed a significant preference for the pair which is embedded in a larger matching structure. This suggests that systematicity does indeed provide a selection filter on which lower-order relations are included in the interpretation of an analogy. Thus, we can go beyond simply postulating a relational preference and add systematicity as a further informational constraint on the interpretation of analogy.

Finally, there is computational support for the structure-mapping account. A computer simulation of the theory, called the Structure-Mapping Engine (SME) (written by Brian Falkenhainer and Ken Forbus), produces psychologically plausible interpretations of analogies and relational metaphors. (Falkenhainer *et al.*, 1986 in press; Gentner, 1988; Gentner *et al.*, 1987).

Given predicate calculus representations of two potential analogs, it uses purely structural principles—*one-to-one correspondence*, *structural consistency*, and *systematicity*—to interpret and evaluate an analogy between two situations. It operates by first finding all possible relational identities between base and target; it then assigns to each of these match hypotheses an evaluation, based on the structural closeness of the match and on a kind of local systematicity by which a given pair of matching predicates is assigned a higher evaluation if their parents also match. SME then sweeps these matching pairs into the largest possible sets consistent with the structural constraints laid out previously and computes an overall evaluation. In addition, it hypothesizes *candidate inferences*, new facts about the target domain that are derived by analogy with the base domain. Thus, SME simulates both the *matching* of existing predicates in the two domains and the *carryover* of hypothesized predicates from one domain to the other. We have compared the performance of SME with that of human subjects for a set of simple analogies using pairs of stories (Skorstad, Falkenhainer & Gentner, 1987). We find that SME's structural evaluations match fairly well with human soundness ratings for story analogies.

Some aspects of structure mapping have received convergent support. There is widespread agreement on the basic elements of one-to-one mappings of objects and structural consistency during matching and carryover of predicates, and many researchers use systematicity or a variant of it as a selection filter (Burstein, 1983; Hofstadter, 1984; Indurkhya, 1985; Kedar-Cabelli, 1985; Van Lehn & Brown, 1980; Winston, 1980, 1982). More generally, the idea that metaphor and analogy involve a mapping of complex knowledge structures is gaining currency in cognitive science (e.g., Black, 1962; Clement, 1987; Gick & Holyoak, 1980, 1983; Hesse, 1966; Hoffman, 1980; Keane, 1988; Kittay, 1987; Lakoff & Johnson, 1980; Miller, 1979; Reed, 1987; Rumelhart & Norman, 1981; Tourangeau & Sternberg, 1982; Verbrugge & McCarrell, 1977).

B. Structural vs. Goal-Driven Models

As discussed previously, many researchers in cognitive science and artificial intelligence have made use of structural principles in modeling analogy. However, some analogy researchers have argued for a stronger focus on contextual goals and plans in addition to (or instead of) structural principles (Burstein, 1983; Burstein and Adelson, 1987; Carbonell, 1981, 1983; Holyoak, 1985; Kedar-Cabelli, 1985). For example, Carbonell (1981) proposed that metaphors and analogies are interpreted by means of an *invariance hierarchy* which captures the degree to which people expect different conceptual relations to be preserved in an interpretation. According to this view, when given a new metaphor or analogy people will first seek an interpretation in terms of a *goal-expectation setting*, then try *planning* and *counterplanning strategies*, then *causal structures*, and

so on through 10 categories, with *descriptive properties* and *object identity* at the bottom of the list.

The goal-centered approach to analogy can be divided into three separable claims, not all of which are maintained by any one researcher:[17]

1. Analogies tend often to be about goals and plans; that is, the higher-order relations that govern the analogy are often goal structures because humans often need to reason about such things (e.g., Carbonell, 1983). In this view, the goal-centered approach can be treated as a specialization of the structural account that attempts to take into account the relative frequency of different kinds of relational structures in human reasoning. Carbonell's invariance hierarchy could be seen as an instance of this effort.
2. Structural principles are important in analogy but must be augmented by some mechanism for taking into account the current cognitive context, including the goals of the person. One reason for this is that structural principles alone are felt to be insufficient to decide among alternative possible interpretations of an analogy (Burstein, 1983; Kedar-Cabelli, 1985). These accounts supplement structural principles with contextual–pragmatic considerations (often in the initial selection of the knowledge given to the matching process) and are, in general, compatible with the views presented here (See Gentner, 1987, 1988a).
3. The strongest version of the goal-centered claim is that the interpretation mechanisms for analogy are defined with respect to the user's goals (e.g, Holyoak, 1985). Thus, an analogy can only be comprehended in the context of a current goal or plan. This strong form of the goal-oriented account cannot be viewed as a specialization of structure mapping. Rather, it seeks to *replace* structural principles with goal-driven selection mechanisms. Since this proposal is a distinct alternative to structure mapping, we discuss it here.

The most extreme of the goal-centered accounts is that given by Holyoak (1985). He proposes that analogy must be modeled as part of a goal-driven processing system: "Within the pragmatic framework, the structure of analogy is closely tied to the mechanisms by which analogies are actually used by the cognitive system to achieve its goals." (Holyoak, 1985, p. 76). He argues that the structure-mapping approach is "doomed to failure" because it fails to take account of goals. In Holyoak's pragmatic account, matching is governed entirely by the relevance of the predicates to the current goals of the problem solver. A crucial difference between the

[17]Occasionally a fourth version of the goal-oriented approach is brought forth: the claim that "In analogy the person has a goal to comprehend the analogy." Since this claim is compatible with all of the views discussed, including the pure structural view, we will not discuss it further.

pragmatic account and the structure-mapping account[18] is that in the pragmatic account the distinction between structural commonalities and surface commonalties is based solely on relevance. Holyoak's definitions of these terms are as follows:

> It is possible, based on the taxonomy of mapping relations discussed earlier, to draw a distinction between *surface* and *structural* similarities and dissimilarities. An identity between two problem situations that plays no causal role in determining the possible solutions to one or the other analog constitutes a surface similarity. Similarly, a structure-preserving difference, as defined earlier, constitutes a surface dissimilarity. In contrast, identities that influence goal attainment constitute structural similarities, and structure-violating differences constitute structural dissimilarities. Note that the distinction between surface and structural similarities, as used here, hinges on the relevance of the property in question to attainment of a successful solution. The distinction thus crucially depends on the goal of the problem solver. (Holyoak, 1985, p. 81)

The key point here is the final sentence: in the pragmatic account, the distinction between surface and structural similarities "hinges on the relevance of the property in question to attainment of a successful solution. The distinction thus crucially depends on the goal of the problem solver." *Structural similarities* are defined as "identities that influence goal attainment," and a surface similarity as "an identity between two problem situations that plays no causal role in determining the possible solutions to one or the other analog." This means that the distinction between surface and structural information—and, therefore, the decision as to what to include in the interpretation of an analogy—must be made with respect to the person's current goals.

Holyoak's emphasis on plans and goals has some appealing features. This account promises to replace the abstract formalisms of structural approach with an ecologically motivated account centered around what matters to the individual. Further, to apply structure-mapping to a problem-solving situation requires at least two kinds of selection criteria: structural selection filters within the matcher, as discussed in this article, and a goal-relevance check on the output of the analogical match process (which is modeled as external to the matching process).[19] Holyoak's account requires only one kind of selection criterion: relevance to the goal of the problem-solver. But there are severe costs to this simplification. First, since structural matches are defined only by their relevance to a set of goals, the pragmatic account requires a context that specifies what is relevant before

[18]It must be mentioned that Holyoak may have modified his views. In a recent talk, Holyoak and Thagard (1987) adopted a more structural approach.

[19]In fact, I have argued elsewhere that there are *three* separate criteria that must be applied to an analogy interpretation: structural soundnes, goal relevance, and validity in the target (Gentner, 1988).

it can operate. Thus, this position fails to capture people's ability to comprehend a new metaphor or analogy without reference to a prior goal context. Such a mechanism could not interpret an analogy in isolation, nor could it interpret an analogy whose point is irrelevant to the current goal context. Such a position clearly fails as an account of the interpretation of analogy. Although prior contextual goals can provide advance expectations that facilitate the interpretation of an analogy, such expectations are not an inherent requirement for computing an interpretation. Indeed, people often perform the reverse computation; they derive the interpretation of an analogy through a structural computation, such as described here, and then infer from this the probable plans and goals of the speaker. For example, in the Clement and Gentner study, subjects derived systematic interpretations of the analogies without requiring any prior problem-solving goals. An example closer to hand is the analogies and metaphors in Table I. Since their meanings are not supported by a current goal context, if the interpretation mechanism requires a goal context it should be impossible to comprehend these comparisons. We leave it to the reader to judge whether this is true.

To be fair, it should be noted that Holyoak's (1985) pragmatic account is meant to apply only to analogies in problem solving. As we have shown, when it is applied to the more general question of how analogy is interpreted it encounters serious difficulties. One could, perhaps, preserve the pragmatic account by postulating separate mechanisms for analogy in a goal context and analogy in isolation. In the former case, "structural commonalities" would be defined in terms of goal relevance; in the latter, they would be defined in terms of predicate structure. However, such a dichotomy would lead to a loss in generality. It seems more reasonable to assume that the basic mechanisms of analogy and similarity operate across different task contexts. Such a view is compatible with the structural account but not with the pragmatic account.

There are further difficulties with the pragmatic account. Because the interpretation of an analogy is defined in terms of relevance to the initial goals of the problem solver, the pragmatic view does not allow for unexpected outcomes in an analogical match. This means that many creative uses of analogy—such as scientific discovery—are out of bounds. Finally, the pragmatic account lacks any means of capturing the important psychological distinction between an analogy that fails because it is *irrelevant* and an analogy that fails because it is *unsound*. In short, although a good case can be made for the need to *augment* structural considerations with goal-relevant considerations (e.g., Burstein, 1983), the attempt to *replace* structural factors such as systematicity with pragmatic factors like relevance is misguided.

C. Literal Similarity

It is interesting to ask if the effects of higher-order structure operate in literal similarity as well as analogy. Although Tversky's (1977) theory of literal similarity makes no distinctions among predicate types, the structure-mapping framework presented previously (See Fig. 1) postulates that common relational structure (as well as common object attributes) is important in literal similarity. In an ongoing research project with Doug Medin and Rob Goldstone, we find support for this claim: subjects' literal similarity judgments show strong effects of relational structure (Goldstone, Medin, & Gentner, 1988). Moreover, the effects of relational similarity appear to be separable from the effects of attributional similarity. For example, given this triad, most people would agree that the first string is more similar to the second than to the third:

Xo Xn To

But if we add exactly one first-order feature (the letter *o*) to each of these strings, we can change this preference so that most people now choose the third string as most similar to the first:

oXo oXn oTo

Since the same lower-order feature was added to each of the three strings, the change in similarity ordering cannot be accounted for as an additive effect of lower-order features. [Indeed, if the vocabulary were restricted to lower-order features this shift would constitute a violation of Tversky's (1977) independence principle.] Instead, it appears that the shift occurs because people are sensitive to the shared symmetry relation between the first and third string in the second set. Thus, relational structure seems to be an important aspect of ordinary similarity as well as analogy and metaphor.

D. Conclusion

We have examined four alternative accounts of the interpretation rules for analogy and metaphor. These accounts vary in their representational assumptions in whether multidimensional spaces, feature sets, or propositional structures are assumed. We have reviewed the structure-mapping theory, which posits a psychological difference between relations and attributes and between systems of relations and isolated relations. The theory is unique in postulating purely structural computations over higher-order predicate representations; they do not depend on the specific content of

the representations or on a set of prior goals or expectations. The evidence presented supports these claims. These structural distinctions appear to be essential for capturing the informational constraints on the interpretation of analogy and metaphor.

This research suggests that analogies and many classes of metaphors can be viewed as devices for highlighting and carrying over relational structure. Because of this, analogies and metaphors allow us to focus on relational commonalities that would otherwise be difficult to express. Further, beyond their communicative uses, analogies and metaphors have enormous conceptual utility as tools for the extraction of relational structure. They allow us to become aware of potentially important relational structures that are not yet explicitly represented in our conceptual and linguistic system, and which may then be abstracted away from the objects to which they apply. In Russell's words, "It must have required many ages to discover that a brace of pheasants and a couple of days were both instances of the number two." Research in analogy and metaphor may provide a way to understand this achievement.

V. Appendix A: Scoring Propositional Structure

Attributionality and relationality are judgments about the conceptual predicate structure underlying the surface language. In most cases, the form of the surface expression makes it clear whether the underlying predicate is an attribute or a relation. For example, predicates that take two or more objects, such as transitive verbs, were scored as expressing relationships between their arguments, e.g., "*X* hit *Y*" and "*X* likes *Y*." Also, comparatives such as "*X* is larger than *Y*," which express relations between attributes of objects were scored as relations. Adjectives often express single-object attributes. These were scored as object attributes whether or not they were stated as an adjectival proposition, e.g., the proposition "*X* is 10 feet tall" was scored as an object attribute.

For the cases discussed so far, there are clear surface signs of their relational or attributional usage. A more difficult set of cases arises when underlying relations are expressed as surface attributes through a process of abstraction (see Miller, 1979). For example, the adjective *soporific* in *X* is soporific" is stated as though it were a quality of *X*, but in fact it conveys relational information that there exist beings whom *X* puts to sleep. It stands for a set of relational statements like "*X* puts *Y* to sleep," "*X* puts *Z* to sleep," etc. These kinds of terms are both relational, in their underlying meaning, and attributional, in that the person has chosen to express the quality as an attribute. Such abstracted relational adjectives were scored as conveying, in moderate degree, both relational and attributional meaning.

VI. Appendix B: Grammatical Categories Used in Syntactic Scoring

Statement type:

A. Nouns
Ordinary
e.g., It has a very large *hat*.
They both live in small *houses*.

Relational
e.g., It is a *container* for something.
(It could contain something).
They are good *providers* for their young.
(They give something to someone else).

B. Adjectives
Ordinary
e.g., It has *big* feet.
They are both *red*.

Relational
e.g., It is an *edible* lead.
(Someone can eat it.)
Both are *soporific*.
(They put a person to sleep.)

C. Comparative Modifiers
Adjectives and adverbs that compare often with *er* suffix).
e.g., Trees are *taller than* buses.
Both run *faster than* their prey.

D. Prepositions
A word that connects a noun or pronoun to another noun, usually with relative spatial location or direction.
e.g., The book *above* the desk is the one.
Fluid flows *from* the heart *to* the lungs.

E. Adverbs
Modifiers of verbs. (usually with *ly* suffix).
e.g., Both move *swiftly*.
Both think *keenly*.

F. Verbs
Transitive
Verbs that require a direct object. Action passes from the subject across the verb to an object of the verb.
e.g., Jerry *hit* the ball.
Karen *took* the books.

Intransitive
Action terminates with the verb. Intransitive verbs do not require a direct object.
e.g., Mary *pondered*.
The motor *raced*.

Linking
Verbs that join the subject and the direct object or modifier together. Implies that they are equal or similar.
e.g., be (is, am) The weather *is* nasty.
taste The pie *tasted* foul.

Auxiliary
Auxiliary or helping verbs are used with other verbs to express complex ideas from tenses, moods, voices, etc. When recording verbs, auxiliary verbs are combined with the verb phrase as a whole.
e.g., is He *is laughing*.
might She *might do*.

Verb plus particle
A verb combined with a short, invariable part of speech, such as a preposition.
e.g., He *hung up* the phone.
She *worked out* in the gym.

Acknowledgments

This research was supported by the Department of the Navy, Office of Naval Research, under Contract No. N00014-79-C-0338, and under the current Contract No. 00014-85-K0559. The developmental study was supported by the National Institute of Education under Contract Nos. NIE-400-80-0030 and NIE-400-80-0031. We thank Allan Collins, Ken Forbus, Don Gentner, Doug Medin, Andrew Ortony, Mary Jo Rattermann, Bob Schumacher, Ed Smith, Yvette Tenney, and Cecile Toupin for comments on this research. We also thank Judith Block, Philip Kohn, Betsy Perry, Patricia Stuart, Edna Sullivan, Ben Teitelbaum, and especially Monica Olmstead for their help with these studies.

References

Black, M. (1962). *Models and metaphors: Studies in language and philosophy*. Ithaca, New York: Cornell University Press.

Burstein, M. H. (1983). Concept formation by incremental analogical reasoning and debugging. In R. Michalski, J. Carbonell, & T. Mitchell (Eds.), *Machine learning* (Vol. 2, pp. 351–369). Los Altos, California: Morgan Kaufman.

Burstein, M., & Adelson, B. (1987). Analogical learning: Mapping and integrating partial mental models. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*, Seattle, pp. 11–22.

Carbonell, J. G. (1981). A computational model of analogical problem solving. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence Vancouver, British Columbia* pp. 147–152.

Carbonell, J. G. (1983). Derivational analogy in problem solving and knowledge acquisition. *Proceedings of the International Machine Learning Workshop Department of Computer Science, University of Illinois at Urbana-Champaign* pp. 12–18.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (ED.), *Discourse production and comprehension* (pp. 1–40). Norwood, New Jersey: Ablex.

Clement, C. A. (1987). *The representation and use of principles derived from abstractions and examples*. Unpublished manuscript.

Clement, C. A., & Gentner, D. (1988). *Systematicity as a selection constraint in analogical mapping*. Unpublished manuscript.

Collins, A., & Gentner, D. (1983). Multiple models of the evaporation process. *Proceedings of the Fifth Annual Conference of the Cognitive Science Society Rochester, New York* Paper Session No. 6.

Collins, A. M., & Gentner, D. (1987). How people construct mental models. In S. Holland and N. Quinn (Eds.) *Cultural models in language and thought* (pp. 243–265). Cambridge: Cambridge University Press.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, **8**, 240–247.

Conner, K., & Kogan, N. (1980). Topic-vehicle relations in metaphor: The issue of asymmetry. In Honeck & R. Hoffman (Eds.), *Cognition and figurative language* (pp. 283–306). Hillsdale, New Jersey: Erlbaum.

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory*, **7**, 397–417.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1988). *The structure-mapping engine. Proceedings of the American Association for Artificial Intelligence*, Philadelphia, pp. 272–277. Revised version to appear in *Artificial Intelligence*, in press.

Forbus, K., & Gentner, D. (1986). Learning physical domains: Towards a theoretical framework. In R. M. Michalski, J. Carbonell, & T. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. II, pp. 311–348). Los Altos, California: Morgan Kaufmann.

Gentner, D. (1980). *The structure of analogical models in science* (Tech. Rep. No. 4451). Cambridge, Massachusetts: Bolt Beranek & Newman.

Gentner, D. (1982). Are scientific analogies metaphors? In D. Miall (Ed.), *Metaphor: Problems and perspectives* (pp. 106–132). Brighton: Harvester Press.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, **7**, 155–170.

Gentner, D. (1986). *Mechanisms of analogical learning*. Paper presented at the Workshop on Analogy and *Similarity*, Monticello, Illinois; to appear in S. Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning*, in press.

Gentner, D. (1988a). Analogical inference and analogical access. In *Analogica*, pp. 63–88. London: Pitman Publishing Co.

Genter, D. (1988b). Structure-mapping in analogical development: The relational shift. *Child Development*, **59**, 47–59.

Genter, D., Falkenhainer, B., & Skorstad, J. (1987). Metaphor: The good, the bad and the ugly. *Proceedings of the Third Conference on Theoretical Issues in Natural Language Processing, Las Cruces, New Mexico*. Also to appear in D. H. Helman (Ed.), *Analogical reasoning: Perspectives of artificial intelligence, computer science, and philosophy*. Boston: Reidel.

Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 99–129). Hillsdale, New Jersey: Erlbaum.

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

Gentner, D., & Schumacher, R. M. (1986). Use of structure-mapping theory for complex systems. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Atlanta pp. 252–258.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, **12**, 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, **15**, 1–38.

Glucksberg, S. (1980). *Remarks*. Presented at the Symposium on Metaphor as Knowledge, meeting of the American Psychological Association, Montreal.

Goldstone, R., Medin, D., & Gentner, D. (1988, April). Relational similarity and the non-independence of features in similarity judgements. Paper presented at the meeting of the Midwestern Psychology Association, Chicago.

Hesse, M. B. (1966). *Models and analogies in science*. Notre Dame, Indiana: University of Notre Dame Press.

Hoffman, R. R. (1980). Metaphor in science. In R. P. Honeck & R. R. Hoffman (Eds.), *Cognition and figurative language* (pp. 393–423). Hillsdale, New Jersey: Erlbaum.

Hofstadter, D. R. (1984). *The Copycat project: An experiment in nondeterministic and creative analogies* (MIT A.I. Laboratory Memo 755). Cambridge, Massachusetts: MIT Press.

Holyoak, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 1, pp. 59–87). New York: Academic Press.

Holyoak, K., & Thagard, P. (1987). *Computational simulation of analogy*. Paper presented at the meeting of the Society for Philosophy and Psychology, San Diego.

Indurkhya, B. (1985). *Constrained semantic transference: A formal theory of metaphors* (Tech. Rep. No. 85/008). Boston: Boston University, Department of Computer Science.

Keane, M. (1988). *Analogical problem solving*. Chichester, U.K.: Ellis Horwood.

Kedar-Cabelli, S. (1985). Purpose-directed analogy. *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*, Irvine, CA, pp. 150–159.

Kittay, E. (1987). *Metaphor: Its cognitive force and linguistic structure*. Oxford: Clarendon Press.

Kittay, E., & Lehrer, A. (1981). Semantic fields and the structure of metaphor. *Studies in language*, **5,1**, 31–63.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Krumhansl, C. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, **85**, 176–184.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Miller, G. A. (1979). Images and models, similes and metaphors. In A. Ortony (Ed.), *Metaphor and thought* (pp. 202–250). Cambridge: Cambridge University Press.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, Massachusetts: Harvard University Press.

Nagy, W. (1974). *Figurative patterns and the redundancy in lexicon*. Doctoral dissertation, University of California at San Diego.

Norman, D. A., Rumelhart, D. E., & the LNR Research Group. (1975). *Explorations in cognition*. San Francisco: Freeman.

Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, **86**, 161–180.

Ortony, A., Vondruska, R. J., Foss, M. A., & Jones, L. E. (1985). Salience, similies, and the asymmetry of similarity. *Journal of Memory and Language*, **24**, 569–594.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). Hillsdale, New Jersey: Erlbaum.

Palmer, S. E., & Kimchi, R. (1985). The information processing approach to cognition. In T. Knapp & L. C. Robertson (Eds.), *Approaches to cognition: Contrasts and controversies* (pp. 37–77). Hillsdale, New Jersey: Erlbaum.

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

Reed, S. K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **13**, 124–139.

Reynolds, R. E., & Ortony, A. (1980). Some issues in the measurement of children's comprehension of metaphorical language. *Child Development*, **51**, 1110–1119.

Rips, L. J., & Turnbull, W. (1980). How big is big? Relative and absolute properties in memory. *Cognition*, **8**, 175–185.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York: Academic Press.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, **7**, 532–547.

Rumelhart, D. E., & Abrahamson, A. A. A. (1973). A model for analogical reasoning. *Cognitive Psychology*, **5**, 1–28.

Rumelhart, D. E., & Norman, D. A. (1981). Analogical processes in learning. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 335–359). Hillsdale, New Jersey: Erlbaum.

Rumelhart, D. E., & Ortony, A. (1977). Representation of knowledge. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, New Jersey: Erlbaum.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, New Jersey: Erlbaum.

Schumacher, R. M., & Gentner, D. (1988). Transfer of training as analogical mapping. *IEEE Transactions on Systems, Man, and Cybernetics*.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, **39**, 373–421.

Skorstad, J., Falkenhainer, B., & Gentner, D. (1987). Anological Processing: A simulation and empirical corroboration. *Proceedings of the Meeting of the American Association of Artificial Intelligence*, Seattle.

Tourangeau, R., & Sternberg, R. J. (1981). Aptness in metaphor. *Cognitive Psychology*, **13**, 27–55.

Tourangeau R., & Sternberg, R. J. (1982). Understanding and appreciating metaphors. *Cognition*, **11**, 203–244.

Tversky, A. (1977). Features of similarity. *Psychological Review*, **84**, 327–352.

Van Lehn, K., & Brown, J. S. (1980). Planning nets: A representation for formalizing analogies and semantic models of procedural skills. In R. E. Snow, P. A. Federico, & W. E. Montague (Eds.), *Aptitude, learning and instruction: Cognitive process analyses* (Vol. 2, pp. 95–137). Hillsdale, New Jersey: Erlbaum.

Verbrugge, R. R., & McCarrell, N. S. (1977). Metaphoric comprehension: Studies in reminding and resembling. *Cognitive Psychology*, **9**, 494–533.

Winston, P. H. (1980). Learning and reasoning by analogy. *Communications of the ACM*, **23**, 689–703.

Winston, P. H. (1982). Learning new principles from precedents and exercises. *Artificial Intelligence*, **19**, 321–350.