

Structure-Mapping: A Theoretical Framework for Analogy*

DEDRE GENTNER

Bolt Beranek and Newman Inc.

A theory of analogy must describe how the meaning of an analogy is derived from the meanings of its parts. In the *structure-mapping theory*, the interpretation rules are characterized as implicit rules for mapping knowledge about a base domain into a target domain. Two important features of the theory are (a) the rules depend only on syntactic properties of the knowledge representation, and not on the specific content of the domains; and (b) the theoretical framework allows analogies to be distinguished cleanly from literal similarity statements, applications of abstractions, and other kinds of comparisons.

Two mapping principles are described: (a) Relations between objects, rather than attributes of objects, are mapped from base to target; and (b) The particular relations mapped are determined by *systematicity*, as defined by the existence of higher-order relations.

When people hear an analogy such as "An electric battery is like a reservoir" how do they derive its meaning? We might suppose that they simply apply their knowledge about reservoirs to batteries, and that the greater the match, the better the analogy. Such a "degree of overlap" approach seems reasonably correct for literal similarity comparisons. In Tversky's (1977) contrast model, the similarity between A and B is greater the greater size of the intersection ($A \cap B$) of their feature sets and the less the size of the two

*This research was supported by the Department of the Navy, Office of Naval Research under Contract No. N00014-79-C-0338, and by the National Institute of Education under Contracts No. NIE-400-80-0030 and NIE-400-80-0031. I thank my colleagues Allan Collins, Ken Forbus, Don Gentner, Ed Smith, and Al Stevens, who collaborated on the development of these ideas; and Susan Carey, John Clement, Andy diSessa, Georges Rey, David Rumelhart, Patrick Winston, and Marianne Wiser for insightful discussions of this approach. I also thank Judith Block, Phillip Kohn, Mary McManamon, Patricia Stuart, Edna Sullivan and Ben Teitelbaum for their help with the research on which this paper is based, and Cindy Hunt for preparing the manuscript. Correspondence and requests for reprints should be sent to Dedre Gentner, Bolt Beranek and Newman Inc., Cambridge, MA 02238.

complement sets ($A - B$) and ($B - A$).¹ However, although the degree-of-overlap model appears to work well for literal similarity comparisons, it does not provide a good account of analogy. The strength of an analogical match does not seem to depend on the overall degree of featural overlap; not all features are equally relevant to the interpretation. Only certain kinds of mismatches count for or against analogies. For example, we could not support the battery-reservoir analogy by remarking (even if true) that batteries and reservoirs both tend to be cylindrical; nor does it weaken the analogy to show that their shapes are different. The essence of the analogy between batteries and reservoirs is that both store potential energy, release that energy to provide power for systems, etc. We can be quite satisfied with the analogy in spite of the fact that the average battery differs from the average reservoir in size, shape, color, and substance.

As another example of the selectiveness of analogical mapping, consider the simple arithmetic analogy $3:6::2:4$. We do not care how many features 3 has in common with 2, nor 6 with 4. It is not the overall number of shared versus nonshared features that counts here, but only the relationship "twice as great as" that holds between 3 and 6 and also between 2 and 4. To underscore the implicit selectiveness of the feature match, note that we do not consider the analogy $3:6::2:4$ better or more apt than the analogy $3:6::200:400$, even though by most accounts 3 has more features in common with 2 than with 200.

A theory based on the mere relative numbers of shared and non-shared predicates cannot provide an adequate account of analogy, nor, therefore, a sufficient basis for a general account of relatedness. In the structure-mapping theory, a simple but powerful distinction is made among predicate types that allows us to state which ones will be mapped. The central idea is that an analogy is an assertion that a relational structure that normally applies in one domain can be applied in another domain. Before laying out the theory, a few preliminaries are necessary.

PRELIMINARY ASSUMPTIONS

1. Domains and situations are psychologically viewed as systems of objects, object-attributes and relations between objects.²

¹According to Tversky (1977), the negative effects of the two complement sets are not equal: for example, if we are asked "How similar is A to B", the set ($B - A$)—features of B not shared by A—counts much more than the set ($A - B$).

²These "objects" may be clear entities (e.g., "rabbit"), component parts of a larger object (e.g., "rabbit's ear"), or even coherent combinations of smaller units (e.g., "herd of rabbits"); the important point is that they function as wholes at a given level of organization.

2. Knowledge is represented here as propositional networks of nodes and predicates (cf. Miller & Johnson-Laird, 1976; Norman, Rumelhart, & the LNR Group, 1975; Rumelhart & Ortony, 1977; Schank & Abelson, 1977). The nodes represent concepts treated as wholes; the predicates applied to the nodes express propositions about the concepts.
3. Two essentially syntactic distinctions among predicate types will be important. The first distinction is between object attributes and relationships. This distinction can be made explicit in the predicate structure: *Attributes* are predicates taking one argument, and *relations* are predicates taking two or more arguments. For example, COLLIDE (x,y) is a relation, while LARGE (x) is an attribute.³
The second important syntactic distinction is between first-order predicates (taking objects as arguments) and second- and higher-order predicates (taking propositions as arguments). 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.
4. These representations, including the distinctions between different kinds of predicates, are intended to reflect the way people construe a situation, rather than what is logically possible.⁴

STRUCTURE-MAPPING: INTERPRETATION RULES FOR ANALOGY

The analogy "A T is (like) a B" defines a mapping from B to T. T will be called the *target*, since it is the domain being explicated. B will be called the *base*, since it is the domain that serves as a source of knowledge. Suppose that the representation of the base domain B can be stated in terms of object

³One clarification is important here. Many attributive predicates implicitly invoke comparisons between the value of their object and some standard value on the dimension. LARGE (x) implicitly means "X is large for its class." For example, a large star is of a different size than a large mouse. But if LARGE (x) is implicitly interpreted as LARGER THAN (X, prototype-x), this might suggest that many surface attributes are implicitly two-place predicates. The theory assumes that only relations that apply *within* the domain of discourse are psychologically stored and processed as true relations. Thus, in the domain of the solar system, a relation such as LARGER THAN (sun, planet), that applies between two objects in the domain, is processed as a relation; whereas an external attributive comparison, such as LARGER THAN (sun, prototype-star), is processed as an attribute.

⁴Logically, a relation R(a,b,c) can perfectly well be represented as Q(x), where Q(x) is true just in case R(a,b,c) is true. Psychologically, the representation must be chosen to model the way people think about the domain..

nodes b_1, b_2, \dots, b_n and predicates such as A, R, R' , and that the *target* domain has object nodes t_1, t_2, \dots, t_m .⁵ The analogy maps the object nodes of B onto the object nodes of T:

$$M: b_i \dashrightarrow t_i$$

These object correspondences are used to generate the candidate set of inferences in the target domain. Predicates from B are carried across⁶ to T, using the node substitutions dictated by the object correspondences.

The mapping rules are

1. Discard attributes of objects:

$$A(b_i) \dashrightarrow A(t_i)$$

2. Try to preserve relations between objects:

$$R(b_i, b_j) \dashrightarrow R(t_i, t_j)$$

3. (The Systematicity Principle) To decide *which* relations are preserved, choose systems of relations:

$$R'(R_1(b_i, b_j), R_2(b_k, b_l)) \dashrightarrow$$

$$[R'(R_1(t_i, t_j), R_2(t_k, t_l))]$$

Higher-order relations play an important role in analogy, as is discussed below.

Notice that this discussion has been purely structural; the distinctions invoked rely only on the syntax of the knowledge representation, not on the content. The *content* of the relations may be static spatial information, as in UNDER(x,y), or FULL(CONTAINER, WATER); or constraint information, as in PROPORTIONAL [(PRESSURE(liquid, source, goal), FLOW-RATE (liquid, source, goal))]; or dynamic causal information, as in CAUSE {AND [PUNCTURE (CONTAINER), FULL(CONTAINER, WATER)], FLOW-FROM (WATER, CONTAINER)}.

⁵Most explanatory analogies are 1-1 mappings, in which $m=n$. However, there are exceptions (Gentner, 1982).

⁶The assumption that predicates are brought across as *identical* matches is crucial to the clarity of this discussion. The position that predicates need only be similar between the base and the domain (e.g., Hesse, 1966; Ortony, 1979) leads to a problem of infinite regress, with similarity of surface concepts defined in terms of similarity of components, etc. I will assume instead that similarity can be restated as identity among some number of component predicates.

KINDS OF DOMAIN COMPARISONS

In the structure-mapping framework, the interpretation rules for analogy can be distinguished from those for other kinds of domain comparisons. The syntactic type of the shared versus nonshared predicates determines whether a given comparison is thought of as analogy, as literal similarity, or as the application of an abstraction.

In this section, different kinds of domain comparisons are described, using the solar system as a common theme. The top half of Figure 1 shows a partial representation of what might be a person's knowledge of our solar system. Both object-attributes, such as YELLOW (sun), and relations between objects, such as REVOLVE AROUND (planet, sun) are shown. Assuming that the hearer has the correct object correspondences, the question is which predicates will be mapped for each type of comparison.

- (1) A *literal similarity* statement is a comparison in which a large number of predicates is mapped from base to target, relative to the number of nonmapped predicates (e.g., Tversky, 1977). The mapped predicates include *both* object-attributes and relational predicates.

EXAMPLE 1. The X12 star system in the Andromeda galaxy is like our solar system.

INTERPRETATION: Intended inferences include both object characteristics—e.g., “The X12 star is YELLOW, MEDIUM-SIZED, etc., like our sun,” and relational characteristics, such as “The X12 planets REVOLVE AROUND the X12 star, as in our system.”

In a literal similarity comparison, all or most of the predicates shown would be mapped.

- (2) An *analogy* is a comparison in which relational predicates, but few or no object attributes, can be mapped from base to target.

EXAMPLE 2. The hydrogen atom is like our solar system.

INTERPRETATION: Intended inferences concern chiefly the relational structure: e.g., “The electron REVOLVES AROUND the nucleus, just as the planets REVOLVE AROUND the sun,” but not “The nucleus is YELLOW, MASSIVE, etc., like the sun.” The bottom half of Figure 1 shows these mapped relations. If higher-order relations are present in the base, they can be mapped as well: e.g., The hearer might map “The fact that the nucleus ATTRACTS the electron CAUSES the electron to REVOLVE around the nucleus” from “The fact that the sun ATTRACTS the planets CAUSES the planets to REVOLVE AROUND the sun.” (This relation is not shown in Figure 1.)

- (3) An *abstraction* is a comparison in which the base domain is an abstract relational structure. Such a structure would resemble Figure 1, except that the object nodes would be generalized physical enti-

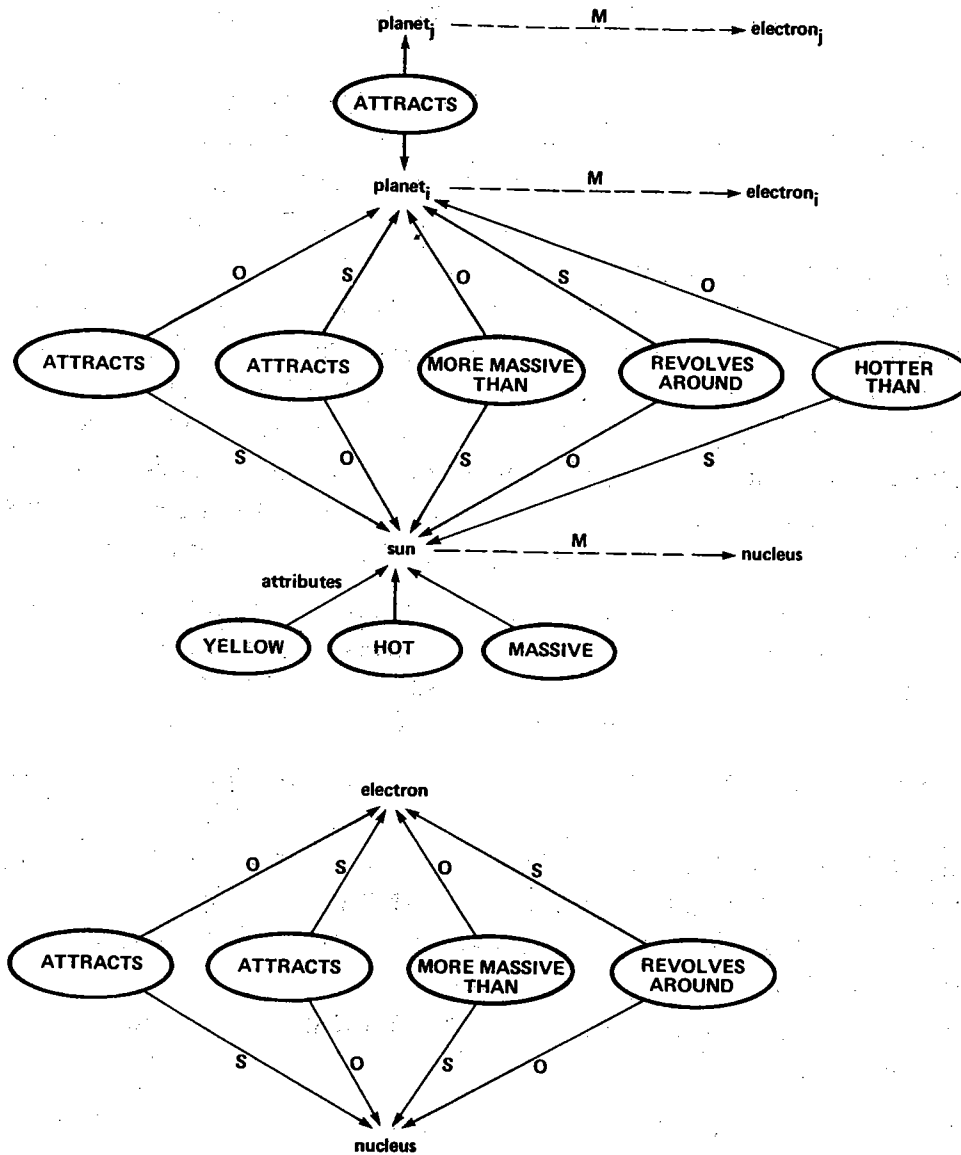


Figure 1. Structure-mapping for the Rutherford analogy: "The atom is like the solar system."

ties, rather than particular objects like "sun" and "planet". Predicates from the abstract base domain are mapped into the target domain; there are no nonmapped predicates.

EXAMPLE 3. The hydrogen atom is a central force system. **INTERPRETATION:** Intended inferences include "The nucleus ATTRACTS the electron;" "The electron REVOLVES AROUND the nucleus." These are mapped from base propositions such as "The central object ATTRACTS the peripheral object;" or "The less massive object REVOLVES AROUND the more massive object." These intended inferences resemble those for the analogy (Example 2). The difference is that in the

