Generative Analogies as Mental Models

Proceedings of the Annual Meeting of the Cognitive Science Society, 3(0)

Gentner, Dedre

1981

Generative Analogies as Mental Models

Dedre Gentner

> Subject explaining electric current: If you increase resistance in the circuit, the current slows down. Now that's like a high--cars on a highway where you--if you notice as you close down a lane, you have cars moving along. Okay, as you go down into the thing, the cars move slower through that narrow point.

When people are reasoning about an unfamiliar domain, they often appear to use analogies, as in the above example from a protocol. Analogies are also used in teaching, as in the following excerpt:

> The idea that electricity flows as water does is a good analogy. Picture the wires as pipes carrying water (electrons). Your wall plug is a high-pressure source which you can tap simply by inserting a plug. The plug has two prongs--one to take the flow to the lamp, radio, or air conditioner, the second to conduct the flow back to the wall. A valve (switch) is used to start or stop flow.

If implicit or explicit analogies are an important determinant of the way people think about complex systems, then it becomes crucial to know exactly how such analogies work. This paper considers the psychological role of these analogies in structuring the target domain.

The first question that must be posed is whether such seeming analogical models do in fact strongly affect the person's conceptualization of the target domain (the Generative Analogy hypothesis), or whether they are merely convenient ways of talking about the domain (the Mere Terminology hypothesis). The mere use of terms borrowed from a given domain--as, for example, when electricity is discussed in terms of moving vehicles or flowing water--is not in itself proof that the speaker is conceiving of electricity as deeply analogous to traffic or to water flow.

To demonstrate that an analogy has generative conceptual power, we must show that nontrivial inferences specific to the base occur in the target. These inferences must be such that they cannot be attributed to shallow lexical associations; e.g. it is not enough to find that the person who speaks of electricity as "flowing" also uses terms such as "capacity" or "pressure". Such usage is certainly suggestive of a Generative Analogy, but it could also occur under the Mere Terminology hypothesis.

The goal here is to show that, at least some of the time, the Generative Analogy hypothesis holds: that deep, indirect inferences in the target follow from use of a given base domain as an analogical model. To do this, we must first decide what inferences should follow from use of a given analogy, and then observe whether the analogies people adopt appear to affect the set of inferences they readily make.

The plan of this paper is (1) to propose a structure-mapping theory of analogy that will allow us to predict the set of inferences that should follow from use of a given analogy; (2) contrast two analogical models for the domain and show that they lead to different indirect inferences; (3) to show that people's inferences concerning simple circuits vary according to which of these models they use; and finally (4) to discuss the general issue of analogical models and structure-mapping.

A structure-mapping theory of analogy. The claim here is that analogies select certain aspects of existing knowledge, and that this selected knowledge can be structurally characterized. First, let's consider what an analogy is not. An analogy such as

(1) An electric circuit with a battery and resistor much like a plumbing system with a reservoir and a constricted section of pipe.

clearly does not convey that all of one's knowledge about the plumbing system should be attributed to the circuit. The inheritance of characteristics is only partial. This might suggest that an analogy is a weak similarity statement, conveying that some but not all of the characteristics of the base system apply to the target system. But this weak characterization fails to capture the distinction between literal similarity and analogical relatedness. Contrast statement (1) with a literal similarity statement like

(2) A hose is like a pipe.

The literal similarity statement (2) conveys that the pipe and the hose share object attributes--e.g. cylindrical shape--as well as sharing similar relationships with other objects--e.g. CONVEY (hose, water)/CONVEY (pipe, water). Statement (1) also conveys considerable overlap in functional relations: e.g. IMPEDE (resistor, current) /IMPEDE (constriction, water). However, it does not convey overlap of objects and their attributes. The resistor as a separate object need not have any qualities in common with the constriction. The analogy, in short, conveys overlap in the system of relations among objects, but no particular overlap in the characteristics of the objects themselves. The literal similarity statement conveys overlap both in relations among the objects and in the attributes of the individual objects.

The analogical models used in science can be characterized as structure-mappings between complex systems. In these analogies, the objects of the known domain, the base domain, are mapped onto the objects of the domain of inquiry, the target domain; the predicates of the base domain--particularly the relations that hold among the nodes--are then applied in the target domain. Structure-mapping analogy asserts that identical operations and relationships hold among nonidentical things. The relational structure is preserved, but not the objects.

Given a particular propositional representation of knowledge we can proceed with an explicit characterization of analogical mapping. A structure-mapping analogy between a target system T and a base system B

This research was supported by the Office of Naval Research and was carried out at Bolt Beranek and Newman and at U.C.S.D. Donald Gentner collaborated on all aspects of this work, but particularly in developing the two analogical models of electricity. I thank Allan Collins, Ken Forbus, Ed Smith and Al Stevens for many insightful comments on this research. Please address correspondence to Dedre Gentner, Bolt Beranek & Newman, 50 Moulton Street, Cambridge, MA 02138.

is an assertion that

1. Given a decomposition of the base and target domains into object nodes $b_1, b_2, \ldots, b_n$ of the base system B and object nodes $t_1, t_2, \ldots, t_m$ of the target system T,
2. The analogical mapping M maps the nodes of B $M: b_i \rightarrow t_i$ into the nodes of T.
3. Then predicates valid in B can be applied in T, using the node substitutions dictated by the mapping:

$$M: [F(b_i,b_j)] \rightarrow [F(t_i,t_j)]$$

Further, the probability that the derived proposition is valid in the target T is greater for relational predicates than for attributes, and greater for high-order relations than for lower-order relations. The strength of the analogical predication increases as we move down the list.

(i) $M: A(b_i) \rightarrow A(t_i)$

(ii) $M: [F(b_i,b_j)] \rightarrow [F(t_i,t_j)]$

(iii) $M: [G(F_1(b_i,b_j), F_2(b_i,b_j)] \longrightarrow [G(F_1(t_i,t_j), F_2(t_i,t_j)]$

Thus, TRUE $[A(b_i)]$ does not strongly suggest TRUE $[A(t_i)]$. Attributional predicates (i) are less likely to carry over than relational predicates (ii); and lower-order relations less likely than higher-order relations (iii).

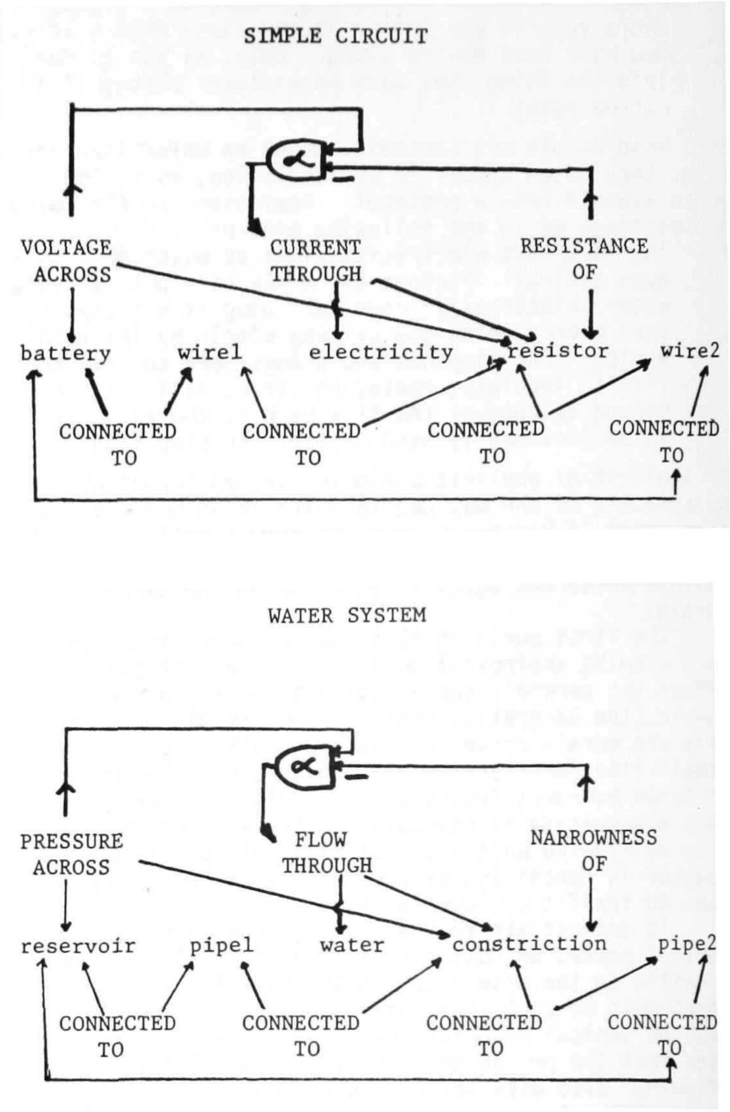
Two models of simple electric circuits. One common analogy used to teach simple electricity is based on plumbing systems. Figure 1 shows the structure-mapping conveyed by this analogy. The object-nodes of the hydraulic base domain (e.g. the reservoir and constriction) are mapped onto the object-nodes (the battery and resistor) of the circuit. Given this correspondence of nodes, the analogy conveys that the relationships that hold between the objects and object-attributes of the hydraulic system also hold between the nodes of the electric system:for example, that current increases with voltage just as rate of water flow increases with pressure; and that current decreases with resistance just as the rate of water flow decreases with degree of constriction.

A second kind of analogy for electric circuits is based on objects moving through chutes. Current is seen as a moving crowd of small objects: voltage is the forward pressure or pushiness of the objects. Like the plumbing model, the moving-object model provides relations that map usefully into the electrical system: If we imagine a source of pushiness corresponding to the battery, and gates in the chute corresponding to the resistors, then the more pushiness, the higher the rate of aggregate motion; the narrower the gates, the lower the rate of aggregate motion.

Although these two analogies convey many of the same relations, in some respects they differ in the aptness of the relational match with the target domain, particularly if we consider slightly more complex circuits (see Gentner and Gentner in press.) The flowing analogy is particularly apt for combinations of batteries, while the moving-object model is superior for combinations of resistors.

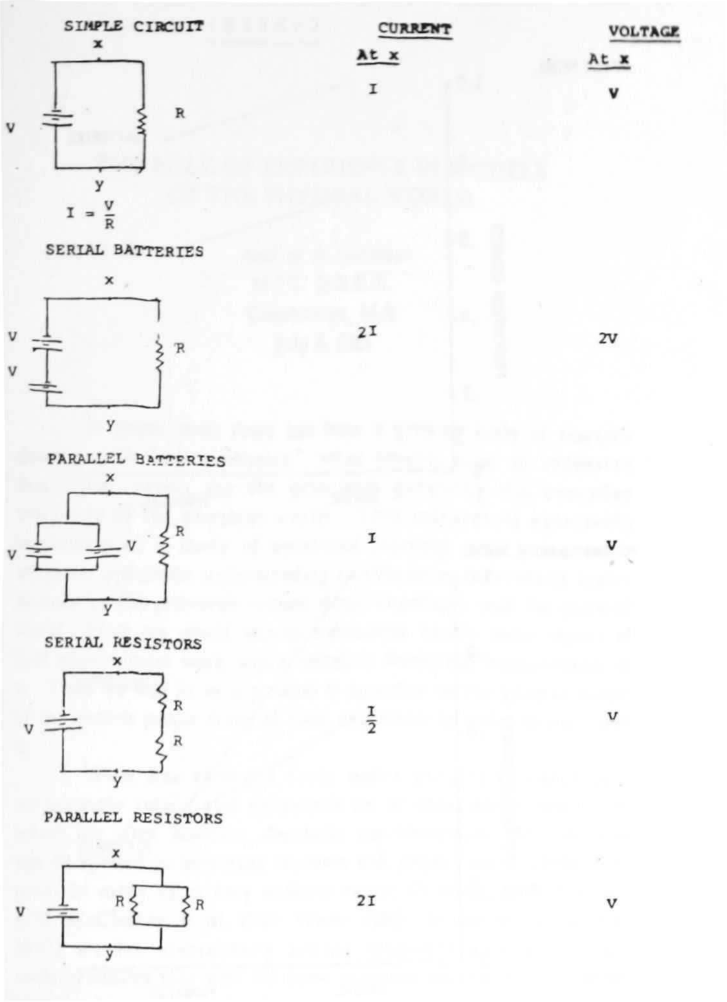
Figure 1

Structural representations of water flow and of simple electric circuit, showing structural overlap



Combinational problems. One way to observe deep indirect inferences in the target domain is to ask about different combinations of components. For example, we can ask how the current in a circuit with two resistors in series or in parallel compares with that in a simple one-resistor circuit. The answers are not obvious. The four circuits generated by series and parallel combinations of batteries and resistors are non-transparent. They provide an excellent way to observe true inferences, as opposed to shallow verbal associations. To deduce the current in these four circuits, the person must move beyond the first-stage naive model of circuitry (shown in Figure 1). This first level of insight is that batteries make for more current, resistors make for less current. These rules hold for batteries and resistors in series, but not for parallel combinations, as shown in Figure 2. Parallel batteries give the same current as a single system, not more; while parallel resistors allow more current than in a simple circuit, not less.

Figure 2.
Current and voltage in simple serial and parallel configuration circuits.



Do analogies make a difference. If subjects are really using their analogical models to understand electronics they should draw on their knowledge of combinational relations among the corresponding components in their respective base domains. Even though the complex circuit problems are couched purely in terms of electronics, we should see differences in subjects' predictions depending on which model they use.

For example, here are two sections of a protocol of a subject trying to predict the current in a parallel-resistor circuit. In the first section, she uses a hydraulics model with reservoirs for batteries, which was her initial model, and derives the wrong answer of *less* current. In the second section, she uses a crowd model suggested to her by the experimenter to derive the correct answer of *more* current:

HYDRAULICS MODEL WITH RESERVOIR

> We started off as one pipe, but then we split into two. Now does that make any difference? I guess it seems to me that this does make a difference. So what we have here is one pipe, one sort of line coming off and then we let it go like that for a while. We let it split off. We have a different current in the split-off section, and then we bring it back together. That's a whole different thing. That just functions as one big pipe of some obscure description. So you should not get as much current.

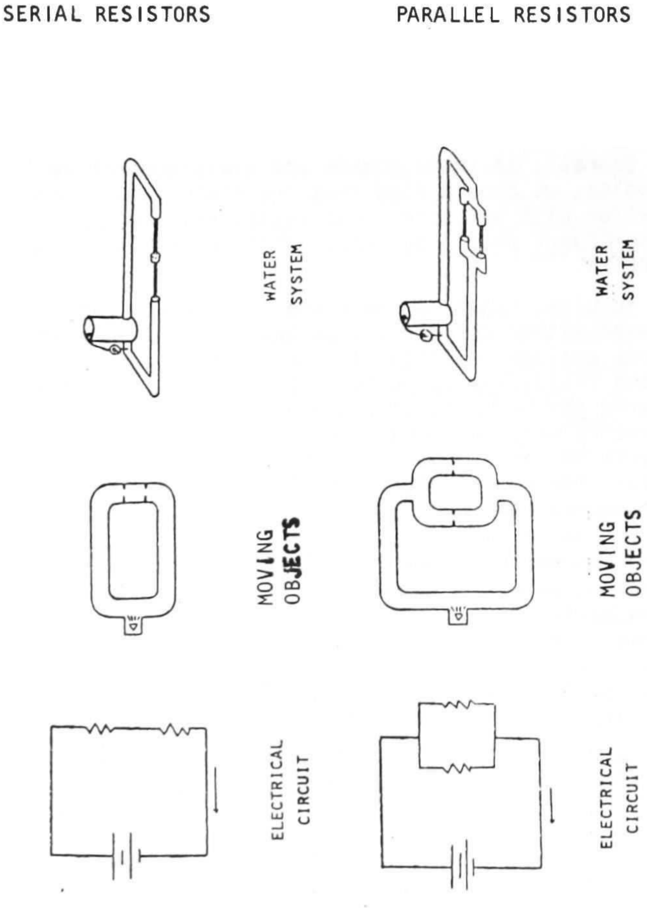
CROWD MODEL

> Again I have all these people coming along here. I have this big area here where people are milling around. I think it is crucial that I separate them though before they get to the gates . . . I can model the two gate system by just putting the two gates right into the arena just like that. So this is one possible model of the two gate system. There are two gates instead of one which seems to imply that the resistance would be half as great if there were only one gate for all those people.

This protocol suggests that models do affect inferences. The following study tests this possibility more on a larger scale. In this study, fairly naive high school and college students were first shown a simple circuit with a battery and a resistor, and then asked to give qualitative solutions for the four combination circuits shown in Figure 2. They were asked to circle whether the current (and voltage) in each of the combination circuits would be greater than, equal to, or less than that of the simple battery-resistor circuit. After they gave their answers for all four combination circuits, they were asked to describe the way they thought about electricity. Then, for each of the four circuit problems, they were asked to circle whether they had thought about flowing fluid, moving objects, or some other way of conceiving of electricity. They were also asked questions about water, to be sure that they understood the base domain.

Figure 3 shows a schematic diagram of parallel and serial resistors in the target and in the two base systems.

Figure 3



Subject who used the fluid model should do well on the battery questions. This is because serial and parallel reservoirs combine in the same manner as serial and parallel batteries, and the combinational distinctions are spatially quite distinct in the water domain. Two equal reservoirs in series (one above the other) give more pressure and hence a greater rate of water flow than a single reservoir. However, since water pressure depends only on height, not on volume, two reservoirs in parallel (side by side) yield only a rate of flow equal to that of a single reservoir. Thus, if the flowing fluid analogy is generative, then subjects with this model (assuming they know the way pressure works in the base) should be able to differentiate serial and parallel configurations of batteries. For resistors, however, the fluid flow model with its constrictions should not, in general, lead to a strong differentiation between serial and parallel resistors. As the protocol above shows, the different between serial and parallel constrictions is fairly opaque in the fluid flow domain; therefore subjects with this model cannot import the correct combinational distinction into the electricity domain. Thus, the prediction is that subjects with the fluid flow model should do better with batteries than with resistors.

For subjects with the moving-objects model, the pattern should be quite different. In this model, configurations of batteries should be relatively difficult to differentiate, since analogies for batteries are hard to find. In contrast, resistors should be better understood. This is because in the moving-objects model, resistors are often seen as gates. Conceiving of resistors as gates should lead to better differentiation between the parallel and serial configurations. If all the objects must pass through two gates one after the other (serial) then the rate of flow should be lower than for just one gate. On the other hand, if the flow splits and moves through two parallel gates, then the rate of flow should be twice the rate for a single gate. Thus subjects using this model should correctly respond that parallel resistors give more current than a single resistor; and serial resistors, less.

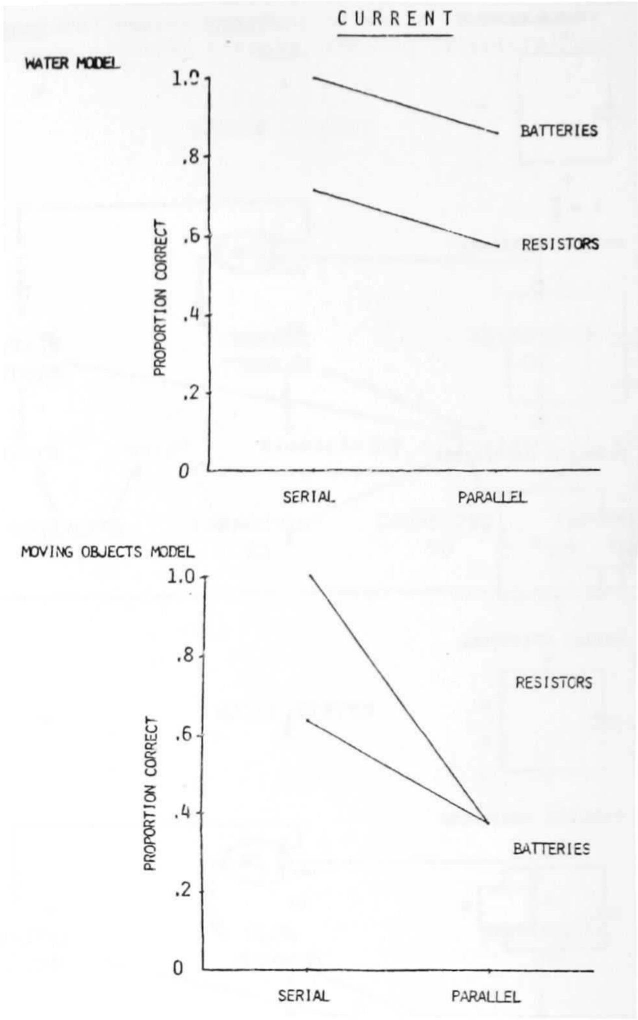
Overall, if these models are truly generative analogies, we should find that the fluid-flow people do better with batteries than resistors, and the moving-object people do better with resistors than with batteries.

Results. Figure 4 shows the results for subjects who used either the fluid-flow analogy or the moving-objects analogy consistently, on all four problems. For the fluid-flow subjects, only those who correctly answered the latter questions about the behavior of reservoirs were included. This was to insure that subjects possessed the requisite knowledge in the base domain. There were nine fluid-flow subjects and seven moving-object subjects.

The patterns of combinational inference are different depending on which model the subject had. As predicted, people who used the fluid-flow model performed better on batteries than on resistors. The reverse is true for the moving-object people. In a Model type x Component type x Topology analysis of variance, the interaction between model type and circuit component is significant; $F(1,13) = 4.53$; $p < .05$.

Conclusions. The results of the study indicate that, for our subjects, the analogies used for electricity were truly generative. Use of different analogies led to systematic differences in the patterns of correct and incorrect inferences in the target domain. Moreover, these combinatorial differences are not easily attributable to shallow verbal associations and communicative patterns. These analogies seem to be truly generative for our subjects; structural relations from the base domain are mapped into the target domain, where they genuinely affect the person's conceptual view of the domain.

Figure 4
Proportion correct on different kinds of circuits for subjects using different models of electricity



The structure-mapping interpretation of the process avoids two extreme positions that often arise in discussions of analogy as explanation: the "vague metaphoricizing" position, which holds that analogy is inherently illogical and unhelpful, and the "appropriate abstractions" position, which emphasizes the fact that analogies convey correct knowledge about the target domain. The structure-mapping view is neutral with respect to whether analogy per se is helpful or harmful. According to the structure-mapping view, the inferences conveyed by a given analogy are not necessarily either correct or incorrect; they are predictable from the predicate structure in the two domains. Relational predicates, particularly those that participate in higher-order systems of relations, are most likely to be mapped; but these may be either correct (as in the case of the moving-object model applied to parallel resistors) or incorrect or indeterminant (as when the moving-object model is applied to batteries). The more we know about the structure of analogy, the better we can design good educational analogies and predict the problem that will occur in use of any given analogy.

References

Gentner, D. and Gentner, D.R. Flowing waters or teeming crowds: Mental models of electric circuits. In D. Gentner & A. Stevens (Eds.) Mental models. Hillsdale, New Jersey: Lawrence Erlbaum Associates in press.