In G.H. Bower (Ed.), The Psychology of Learning and Motivation, Vol. 19 (pp. 59-87). New York: Academic Press.

# THE PRAGMATICS OF ANALOGICAL TRANSFER

*Keith J. Holyoak**

UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN



## I. Introduction

One of the peaks of current achievement in artificial intelligence is the design of expert systems for problem solving. A typical system is provided with detailed knowledge about a specific domain, such as potential defects of an electronic device, and then uses this knowledge in conjunction with inference procedures to mimic the skill of an expert troubleshooter. There are reasons to suspect, however, that the current generation of expert systems is approaching an upper limit in efficacy. The extreme domain specificity of the knowledge incorporated in expert systems leads to what Holland (1984) terms "brittleness"—small changes in the domain to which the system is to be applied typically require extensive human intervention to redesign the knowledge incorporated into the system.

In psychological terms, brittleness reflects a lack of capacity to transfer knowledge based on past experience to novel situations. In fact, current expert systems

*Present address: Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.




ISBN 0-12-543319-0

typically learn nothing at all from experience in solving problems. New knowledge is only acquired as the result of its implantation by the human programmer. In contrast, one of the hallmarks of human intelligence (as well as that of other higher animals) is that the knowledge of the system is self-amplifying; that is, current knowledge, under the guidance of various forms of environmental feedback, is used to modify and augment the knowledge store in an adaptive fashion. Furthermore, learning is possible despite various sources of uncertainty about the state of the environment or the consequences of possible actions that the organism might take.

Following Holland, Holyoak, Nisbett, and Thagard (1985), I will use the term *induction* to refer to all inferential processes that expand knowledge in the face of uncertainty. Developing a descriptive account of human induction may well be the key to providing a prescriptive remedy for the brittleness of systems designed to display artificial intelligence. Many forms of induction involve recombinations and transformations of existing knowledge to generate plausible hypotheses relevant to the organism's interactions with its environment. The focus of the present article will be on one variety of inductive recombination—the use of analogies (particularly analogies between situations drawn from relatively remote knowledge domains) to solve novel problems and to form generalized rules.

My intent is to integrate two lines of empirical and theoretical work in which I have been involved for the past few years. The first is a series of studies of analogical problem solving begun in collaboration with Mary Gick (Gick & Holyoak, 1980, 1983) and continued with others. The second is an overarching pragmatic framework for induction proposed by Holland *et al.* I will attempt to sketch the global view of analogy that emerges from the pragmatic framework as well as specific research issues and hypotheses. The central issues with which I will be concerned involve the need to specify the circumstances under which people are able to notice analogies and put them to appropriate uses. Accordingly, the first part of the article will present a synopsis of relevant aspects of the framework developed by Holland *et al.* This framework will then be used to outline a pragmatic view of analogy. In subsequent sections I will review empirical research, particularly relatively recent work, that explores the impact of various types of similarity on analogical transfer between problems.

## II. A Pragmatic Framework for Induction

### A. The Pragmatic Context of Induction

In a cognitive system with a large knowledge base the set of potential inferences that might be drawn, and the potential recombinations that might be explored, will be indefinitely large. The immediate problem is that without some



basic constraints virtually all the possible extensions of the knowledge base will be utterly useless. For example, if the system knows that snow is white, it might proceed to apply a standard rule of inference in propositional logic to deduce that "either snow is white or rhubarb grows in the Sahara." Such deductions, although assuredly valid, are almost as assuredly pointless. And if unconstrained deduction is a fruitless exercise, unconstrained induction can easily be fatal. If an organism drinks water repeatedly and thereby is led to the confident generalization that all liquids are refreshing, its first encounter with turpentine may have disastrous consequences.

Holland *et al.* argue that progress in understanding induction in philosophy, psychology, and artificial intelligence has been stunted by misguided attempts to specify purely syntactic constraints on induction without attention to the relationship between induction and either the goals of the system or the context in which induction occurs. Indeed, as noted above, the critique extends to traditional accounts of deductive inference as well. (See Cheng & Holyoak, 1985; and Cheng, Holyoak, Nisbett, & Oliver, 1985, for evidence favoring a pragmatic approach to human deductive inference.) In a later section I will illustrate the general critique with respect to a particular syntactic account of analogical transfer.

From the pragmatic perspective, the central problem of induction is to specify processing constraints ensuring that the inferences drawn by a cognitive system will tend to be (a) relevant to the system's goals and (b) plausible. What inductions should be characterized as plausible can only be determined with reference to the current knowledge of the system. Induction is thus highly context dependent, being guided by prior knowledge activated in particular situations that confront the system as it seeks to achieve its goals. The study of induction becomes the study of how knowledge is modified through its use.

Figure 1 sketches the organization of the kind of processing system that places induction in a pragmatic context. The key ideas are that induction is (a) directed by problem-solving activity and (b) based on feedback regarding the success or failure of predictions generated by the system. The current active goals of the system, coupled with an activated subset of the system's current store of knowledge, will provide input to inferential mechanisms that will generate plans and other types of predictions about the behavior of the environment. These predictions, coupled with receptor input (both perceptual representations of the environment and information about internal states of the system), will be fed back to other inferential mechanisms. A comparison of predictions and receptor input will yield information about predictive successes and failures, which will in turn trigger specific types of inductive changes in the knowledge store.[1]

[1]In addition to its role in induction, receptor input can directly alter both the knowledge store (by creating memory traces of perceptual inputs) and the active goals. Constructed plans can also alter goals (by generating new subgoals).

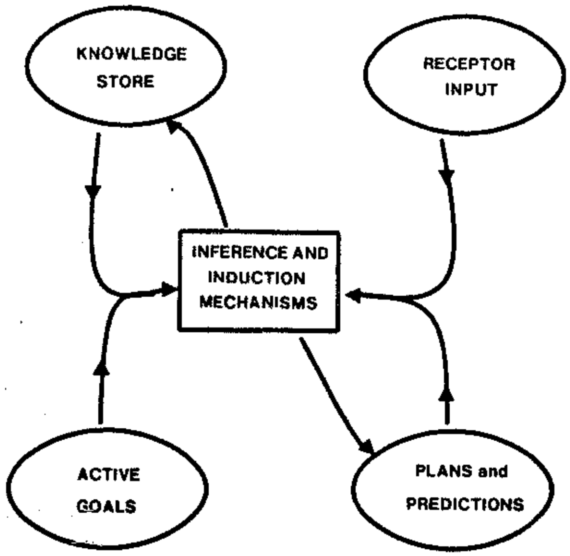


Fig. 1. A model of problem-directed, prediction-based induction.

As Fig. 1 indicates, no strong distinction is made between inferential and inductive mechanisms, since many specific mechanisms play a dual role. As we will see below, analogy provides a clear illustration of this duality of function in that a useful analogy will typically both guide generation of a specific solution plan and trigger induction of a generalized version of the plan.

Within the context of an information-processing system that succeeds in imposing pragmatic constraints on induction, the kind of "inference overload" described earlier is unlikely to arise. Indeed, a person confronted with a problem of an unfamiliar type may suffer not from a surfeit of ready inferences, but rather from lack of even a single plausible approach to the problem. In just this type of problem-solving context an analogy drawn from a different, better understood domain may plan an important role as an inferential and inductive mechanism.

### B. Models and Morphisms

According to the sketch in Fig. 1, the immediate input to inductive mechanisms is the product of a comparison between the system's internally generated predictions and receptor input. At the most general level, the inductive goal of the system is to refine its knowledge store to the point at which its predictions about the environment are sufficiently accurate as to achieve its goals in interacting with the environment. A representation of some portion of the environment that generates predictions about its expected behavior is termed a *mental model*. The function of induction is to refine mental models.



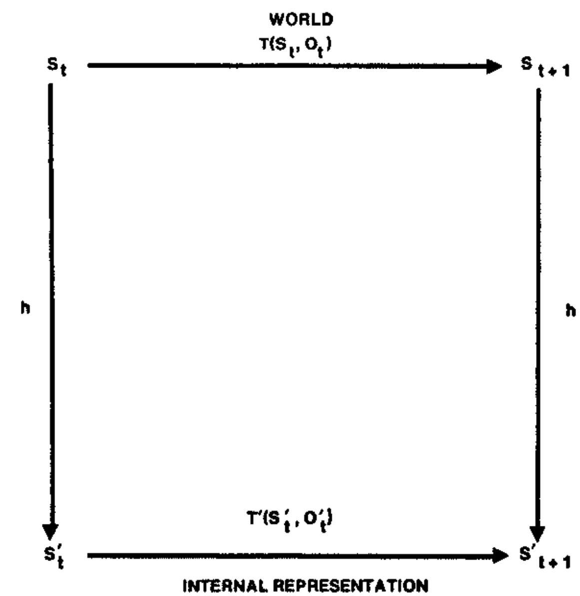


Fig. 2. A mental model structured as a homomorphism.

We can describe the notion of an internal model in terms of *morphisms*, as schematized in Fig. 2. The environment can be characterized in terms of *states* and a *transition function* that relates each state to its successor; that is, the state of the environment at time $t + 1$, $S_{t+1}$, is given by $T(S_t, O_t)$, where $T$ is the transition function and $O_t$ is the output from the cognitive system at time $t$ (i.e., its overt actions). In a mental model, the components of environmental states as well as system outputs are aggregated into categories by a *mapping function* $h_{t+1}$, treating members of each category as indistinguishable for the purposes of the model. For example, many distinct objects in the environment may be categorized as chairs or as people. The model will specify a transition function $T'$ that is intended to mimic sequences of transitions among environmental states up to a pragmatically appropriate level of precision. For example, a fragment of $T'$ might correspond to a rule stating that if a person sits on a chair at time $t$, then the person will be supported at $t + 1$.

If a model constitutes a valid homomorphism, it will have the property of *commutativity;* that is, carrying out a transition in the environment and then determining the category of the resulting state will have the same outcome as determining the category of the initial state and then carrying out the transition in the model. Specifically,

$$h[\mathrm{T}(S_t,\ O_t)] = T'[h(S_t),\ h(O_t)]$$

In a realistic mental model of a complex environment, commutativity will sometimes fail. This will happen in just those circumstances in which the model generates a prediction that does not correspond with subsequent receptor input. Such failures of a model are the triggers for inductive change. One of the most basic types of change is the generation of subcategories corresponding to aberrant cases (e.g., identifying a subtype of the category "bird," the members of which violate the expectation that birds will fly) along with a corresponding refinement in $T'$. For example, an initial expectation that might be stated as, "If a bird is pursued, then it will fly away," might be augmented with a new expectation, "If an ostrich is pursued, then it will run away." Note that the new, more specialized part of the transition function need not replace the more general one. If the system simply prefers more specific expectations in cases of conflict, then an overly general expectation may often serve as a useful default which will occasionally be overridden by a more specific exception. The categories and transitions at each successive level of specificity will further refine the predictive adequacy of the model. The overall structure of the model can be defined recursively as a nested set of morphisms, termed a *quasimorphism*, or *q-morphism*. (For a formal definition, see Holland *et al.*, 1985.) A q-morphism is equivalent to an extended notion of default hierarchy.

Several aspects of the above characterization of mental models as morphisms will prove important for an account of analogical problem solving. First, a solution plan can be viewed as a model in which the initial state is a problem representation, the final state is a representation of the class of goal-satisfying states, and the transition function specifies a plan for transforming the former into the latter. Second, we will see that a valid problem analogy can itself be viewed as a morphism. Furthermore, this characterization is useful in differentiating important from less important differences between analogs. Third, the initial solution plan constructed from an analogy is often imperfect in much the same way as a mental model may in general be imperfect, triggering similar sorts of inductive corrections.

### C. A Rule-Based Representation of Mental Models

The characterization of mental models in terms of morphisms is one step removed from a description of a potential process model. This next step is taken by representing both the mapping function $h$ and the internal transition function $T'$ by *condition–action rules* (i.e., rules of the form "If ⟨condition⟩ then ⟨action⟩."). Models based on condition–action rules are prevalent in both cognitive psychology and artificial intelligence; however, the framework proposed by Holland *et al.* (1985) is most directly derived from a type of parallel rule-based system termed a *classifier system* (Holland, 1975, in press). For the purposes of the present article it will not be necessary to describe the properties of such systems in detail. Only the relationship between types of rules and components of mental models is directly relevant.



The building blocks of mental models are *empirical* rules—rules that describe the environment and its behavior.[2] The rules that constitute the transition function, generating expectations about the behavior of the environment, are termed *diachronic* rules because they describe expected changes in the environment over time. The rules that constitute the mapping function $h$ are termed *synchronic* rules because they have the function of performing atemporal categorizations of the components of environmental states. For example, a typical synchronic rule would be, "If an object is small, feathered, and builds nests in trees, then it is a bird." A diachronic rule might be, "If an object is a bird, and it is chased, then it will fly away."[3]

Synchronic rules capture the kind of categorical and associative information often represented in static semantic networks, while diachronic rules can represent information about the expected effects of system actions, such as problem-solving operators (Newell & Simon, 1972). An advantage of representing both categorical and action-related information in a common rule format is that both are then subject to the same inductive pressures. Similarly, activation of both types of information is determined by the same processing constraints. In the framework proposed by Holland *et al.* (1985), several important principles govern the organization and processing of rules. First, on a single cycle of activity a limited number of rules may operate in parallel. Executed rules post *messages*—pieces of declarative information—that determine the subsequent behavior of the system. Messages may also be generated by receptor inputs and by information retrieved from declarative memory stores. Some messages will specify current goals of the system; those goals will largely direct the activity of the system, and goal attainment will constitute a basic source of "reward" for rules.

On each cycle of processing activity, active messages are matched against the conditions of rules. Those rules that have their conditions fully matched become competing candidates for execution. These candidate rules then place "bids" to determine which of them will be executed. The size of the bid made by each rule is determined by three factors: (a) the *specificity* of the rule's condition (more specific rules make larger bids); (b) the *strength* of the rule (a numerical measure, subject to inductive revision, which indicates the past usefulness of the rule

[2]Holland *et al.* distinguish empirical rules, which directly model the environment, from more general *inferential* rules, which manipulate current information to modify the knowledge store (and hence constitute the "inferential mechanisms" indicated in Fig. 1).

[3]Although time is a critical dimension in problem solving, the synchronic/diachronic distinction can be viewed more generally as one between state re-descriptions versus state transitions, even when the transitions do not correspond to a temporal dimension.

in achieving the system's goals); and (c) the *support* accruing to the rule from the messages that matched it. The latter is a relatively volatile measure of the current activation level of the messages satisfying the rule's condition.

Rules thus compete for the "right" to post messages and hence direct subsequent processing. A number of rules may be jointly executed, allowing the system limited parallelism. The parallel nature of the system allows multiple rules to act together to describe complex (and possibly unique) situations. Inductive mechanisms will favor the development of *clusters* of rules that often work well together. A rule cluster will consist of multiple rules with identical or overlapping conditions. Because the only actions of rules are to post messages, problems of conflict resolution are minimized. Multiple rules can post contradictory messages, and these can coexist until either one message acquires sufficient support to effectively suppress its alternatives, or the need for effector action creates a demand for an unambiguous decision.

As the system operates, the rules will be subject to a variety of inductive pressures. The strengths of existing rules will be revised as a function of their efficacy in attaining goals. In addition, new rules will sometimes be generated in response to particular states of the system, such as failures of predictions, and the new rules will then compete with those already in the system. Analogy, the topic of the present article, is but one of many mechanisms for generation of plausibly useful new rules. (See Holland *et al.*, 1985, for discussion of other inductive mechanisms.)

## III. The Function and Structure of Analogy

Analogy is a broad topic, extending well beyond the domain of problem solving *per se* into the realms of argumentation and literary expression (Holyoak, 1982). In this article I will focus on the role of analogy in problem solving, in keeping with both the theoretical framework outlined above and the scope of our empirical work. In general, analogy is used to generate new rules applicable to a novel *target* domain by transferring knowledge from a *source* domain that is better understood. The overall similarity of the source and target domains can vary enormously along a continuum from the mundane to the metaphorical. At the mundane end of the continuum, it is commonplace for students learning such activities as geometry theorem proving and computer programming to use initial examples as analogical models for solving subsequent problems (Anderson, Greeno, Kline, & Neves, 1981; Pirolli & Anderson, 1985). Our own research has explored the use of problem analogies toward the metaphorical end of the continuum, involving structurally similar situations drawn from superficially dissimilar domains.



### A. An Example: The Convergence Analogy

In order to make our discussion more concrete, let us consider a particular type of analogy we have used in many experiments with college students. The target problem we have used most often is the "radiation problem" first studied by Duncker (1945). The problem runs as follows:

> Suppose you are a doctor faced with a patient who has a malignant tumor in his stomach. It is impossible to operate on the patient, but unless the tumor is destroyed the patient will die. There is a kind of ray that at a sufficiently high intensity can destroy the tumor. Unfortunately, at this intensity the healthy tissue that the rays pass through on the way to the tumor will also be destroyed. At lower intensities the rays are harmless to healthy tissue, but will not affect the tumor either. How can the rays be used to destroy the tumor without injuring the healthy tissue?

This problem is reasonably realistic, since it describes a situation similar to that which actually arises in radiation therapy. On the other hand, it is not a problem that typical college subjects can easily classify as an example of a familiar problem type. In terms of the kind of rule-based system described above, there are no diachronic rules immediately available to construct a transition from the initial problem state to a goal-satisfying state. The problem solver might imagine the possibilities of altering the effects of the rays or altering the sensitivities of the healthy tissue and/or tumor. However, such abstract operators do not specify realizable actions. As a result, the problem is seriously "ill-defined" (Reitman, 1964).

Analogy differs from other inferential mechanisms in that it is less directly focused on the immediate problem situation. To solve a problem by analogy one must attend to information other than the problem at hand. Precisely because few strong rules will be available for directly dealing with an ill-defined problem, weaker synchronic rules that activate associations to the target may have an opportunity to direct processing. Analogy provides a mechanism for augmenting the mental model of an unfamiliar situation with new rules derived from a source analog.

In the case of the radiation problem, an analogy might be used to generate rules that provide more specific operators. This possibility was tested in the initial experiment performed by Gick and Holyoak (1980). The experimenters attempted to demonstrate that variations in the solution to an available source analog can lead subjects to generate qualitatively different solutions to the target. In order to provide subjects with a potential source analog, the experimenters first had subjects read a story about the predicament of a general who wished to capture a fortress located in the center of a country. Many roads radiated outward from the fortress, but these were mined so that although small groups could pass over them safely, any large group would detonate the mines. Yet the general needed to get his entire large army to the fortress in order to launch a successful attack. The general's situation was thus substantially parallel to that of the doctor in the radiation problem.

Different versions of the story described different solutions to the military problem. For example, in one version the general discovered an unguarded road to the fortress and sent his entire army along it; whereas in another version the general divided his men into small groups and dispatched them simultaneously down multiple roads to converge on the fortress. All subjects were then asked to suggest solutions to the radiation problem, using the military story to help them. Those who read the former version were especially likely to suggest sending the rays down an "open passage," such as the esophagus, so as to reach the tumor while avoiding contact with healthy tissue. In contrast, subjects who received the latter story version were especially likely to suggest a "convergence" solution—directing multiple weak rays at the tumor from different directions. Across many comparable experiments, Gick and Holyoak found that about 75% of college students tested generated the convergence solution after receiving the corresponding military story and a hint to apply it. In contrast, only about 10% of students generated this solution in the absence of a source analog, even though most subjects would agree the solution is an effective one once it was described to them. The mapping between the source and target was occasionally revealed in the protocols of subjects who spoke as they worked on the problem:

> Like in the first problem, the inpenetrable fortress, the guy had put bombs all around, and the bombs could be compared to the healthy tissue. And so they had to, they couldn't go in *en masse* through one road, they had to split up so as not to damage the healthy tissue. Because if there's only a little bit of ray it doesn't damage the tissue, but it's all focused on the same spot. (Gick & Holyoak, 1980, p. 327)

## B. Analogical Mapping and Schema Induction

Analogical problem solving involves four basic steps. These are (1) constructing mental representations of the source and the target, (2) selecting the source as a potentially relevant analog to the target, (3) mapping the components of the source and target, and (4) extending the mapping to generate a solution to the target. These steps need not be carried out in a strictly serial order, and they will interact in many ways. For example, a partial mapping with the target is typically required to select an appropriate source. Also, because mapping can be conducted in a hierarchical manner, the process may be iterated at different levels of abstraction. Nonetheless, these four steps impose a useful conceptual organization on the overall process.

The correspondences between the convergence version of the military story and the radiation problem are shown in Table I. Even though the particular objects involved (e.g., army and rays, fortress and tumor) are very different, the basic structural relations that make the convergence solution possible are present in both. The goal, resources (and other objects), operators, and constraint are structurally similar, and hence can be mapped from one problem to the other. Because the military story provides clear problem-solving operators (e.g., "divide the army"), subjects are able to use the mapping to construct corresponding operators (e.g., "reduce ray intensity") that can be used to solve the ray problem.

The abstract structure common to the two problems can be viewed as a *schema* for convergence problems—a representation of the class of problem for which convergence solutions are possible. The convergence schema implicit in the two analogs is sketched at the bottom of Table I. The schema represents an abstract category of which the specific analogs are instances. As we will see below, analogy is closely related to the induction of category schemas by generalization. Indeed, schema induction can be viewed as the final step in analogical transfer.



TABLE I

Correspondences Among Two Convergence Analogs and Their Schema

| |
|---|
| **Military problem** |
| Initial state |
| Goal: Use army to capture fortress |
| Resources: Sufficiently large army |
| Operators: Divide army, move army, attack with army |
| Constraint: Unable to send entire army along one road safely |
| Solution plan: Send small groups along multiple roads |
| Outcome: Fortress captured by army |
| **Radiation problem** |
| Initial state |
| Goal: Use rays to destroy tumor |
| Resources: Sufficiently powerful rays |
| Operators: Reduce ray intensity, move ray source, administer rays |
| Constraint: Unable to administer high-intensity rays from one direction safely |
| Solution plan: Administer low-intensity rays from multiple directions simultaneously |
| Outcome: Tumor destroyed by rays |
| **Convergence schema** |
| Initial state |
| Goal: Use force to overcome a central target |
| Resources: Sufficiently great force |
| Operators: Reduce force intensity, move source of force, apply force |
| Constraint: Unable to apply full force along one path safely |
| Solution plan: Apply weak forces along multiple paths simultaneously |
| Outcome: Central target overcome by force |

Because the information in a problem schema can be represented by a set of interrelated synchronic and diachronic rules, a schema will be represented as a rule cluster of the sort described earlier.

### C. The Pragmatics of Analogy

Within our pragmatic framework for induction, the most basic questions regarding analogy concern the manner in which it can be used to help solve problems. In particular, an account of analogy must address two related puzzles. First, how can a relevant source analog be found efficiently? The target problem will typically be related in some way or another to an enormous range of knowledge, most of which will be entirely unhelpful in generating a solution. Second, once a relevant analog is identified, what determines which of the properties of the source will be used to develop a model of the target problem? Especially when the source and target are highly dissimilar on a surface level, only a small subset of knowledge about the source can be transferred to the target.

In general, a useful source analog will be one that shares multiple, goal-related properties with the target. Goal-related diachronic rules attached to the source analog will provide the basis for the generation of new diachronic rules appropriate to the target problem. An analogy is thus ultimately defined with respect to the system's goals in exploring it. For this reason, as will be argued later, syntactic approaches to analogy, which do not consider the impact of goals on analogical transfer, are doomed to fail.

What does it mean to model a problem by analogy? Analogy involves "second-order" modeling—a model of the target problem is constructed by "modeling the model" after that used in the source problem. As schematized in Fig. 3, the model of the source problem is used as a model of the target problem, generating a new model that can be applied to the novel situation (cf. Fig. 3 with Fig. 2). In Fig. 3, the source model provides a morphism for some aspects of the world (labeled "World A"). A model of the aspects of the world involved in the target problem ("World B") is constructed by means of the analogical mapping (indicated by dark arrows) from the target to the source. As indicated in Fig. 3, in an ideal case the resulting target model will be isomorphic to the source model; that is, goals, objects, and constraints will be mapped in a one-to-one fashion so that corresponding operators preserve the transition function of the source (i.e., the function $T'_B$ in the target model mimics the function $T'_A$ in the source).

Note that even in the ideal case not all elements of the source situation need be mapped, but only those included in the model of the source, that is, those causally relevant to the achieved solution. In any problem model the components are directly relevant to the solution plan: the goal is a *reason* for it; the resources *enable* it; the constraints *prevent* alternative plans; and the outcome is the *result* of executing the solution plan. In terms of rules, the necessary mapping involves



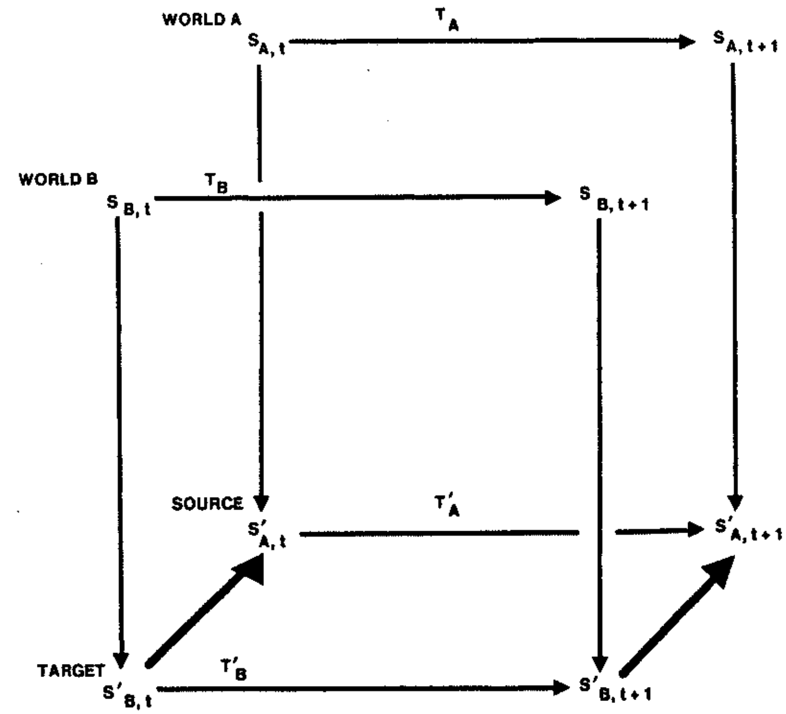


Fig. 3. Analogy as a "second-order" model based on a mapping between models of the source and target domains.

properties of the source included in the conditions of those diachronic rules that constitute the relevant transition function. By defining analogy in terms of relationships between problem models, it is possible to delimit the information transferred from source to target in a principled way. [Hesse (1966) was the first to stress the central importance of causal elements in analogical transfer.]

In practice, however, the initial target model derived by analogy will be less than isomorphic to the source, and even an adequate target model will typically fall short of this ideal. Since the target problem will not be adequately modeled *prior* to the mapping process (otherwise the analogy would be superfluous), the initial mapping with the source will inevitably be partial. Initiation of the mapping process will depend on the problem solver having first identified *some* element of the target problem that is similar to an element of the source model, that is, an element included in either the source model or some generalization thereof. In the case of interdomain analogies the similarities of the source and target are primarily relational. Hence the initial mapping will typically involve detection of an abstract similarity between corresponding goals, constraints, object descriptions, or operators, which constitute the implicit schema common to the two analogs.

Elements of the implicit schema that are identified in the target can serve as retrieval cues to access relevant source analogs as well as to initiate the mapping process once a source analog is available. More will be said later about the retrieval of source analogs. Once the relevance of the source is considered and an initial partial mapping has been established, the analogical model of the target can be developed by extending the mapping. Since models are hierarchically structured and the mapping will usually be initiated at an abstract level, the extension process will typically proceed in a top-down fashion. As the source model is "unpacked," the analogist will attempt to construct a parallel structure in the target model using elements appropriate to the new domain. New rules for the target domain can be generated by substitution of mapped elements in corresponding rules for the source domain.

For example, a subject in one of Gick and Holyoak's experiments might first establish a mapping between the doctor's goal of "using rays to destroy a tumor" and the general's goal of "using an army to capture a fortress," since both are instances of the implicit schema of "using a force to overcome a target." The rays are now (tentatively) mapped with the army; accordingly, the analogist will attempt to construct operators for acting on the rays which match those that act on the army in the source model. For example, since the army could be divided into small groups, it follows that the high-intensity rays could be "divided into small rays" (or, in terms more appropriate to the target domain, "divided into several low-intensity rays"). By substitution of mapped elements, a source rule such as, "If the goal is to get a large army to a central fortress, and a constraint prevents sending the entire army along one road, then divide the army into small groups and send them along multiple roads simultaneously," can be used to generate the critical target rule, "If the goal is to get high-intensity rays to a central tumor, and a constraint prevents administering high-intensity rays from one direction, then divide the high-intensity rays into several low-intensity rays, and administer the low-intensity rays from multiple directions simultaneously."

This process of model development will continue until an adequate target model is created or until a point is reached at which the analogy begins to "break down." What does it mean for an analogy to break down? Holyoak (1984b) distinguishes between four types of mapping relations. *Identities* are those elements that are the same in both analogs (e.g., the generalized goal of "using a force to overcome a target"). The identities are equivalent to the implicit schema. *Indeterminate correspondences* are elements that the analogist has yet to map.

The other two types of mappings involve known differences between the two analogs. For example, the mapped objects of "rays" and "army" obviously will generate a host of differences when the concepts are analyzed. However, differences do not necessarily impair the morphism (after all, the problems are only supposed to be analogous, not identical). *Structure-preserving differences* are those that allow construction of corresponding operators, and hence maintain the transition function of the source model. For example, an army is visible whereas radiation is not; however, since no operators necessary to achieving a solution are blocked simply because the rays are invisible, this difference is structure preserving.



Other differences, however, will be *structure violating* because they prevent the construction of corresponding operators. For example, an army, unlike rays, consists of sentient beings. Accordingly, the general can simply tell his men to "divide up," and the army can be expected to regroup appropriately without further intervention. Rays, of course, do not respond to such cavalier treatment. Although they can in a sense be "divided," the requisite operator will be of a very different type. Subjects often introduce multiple "ray machines" (with no counterparts in the source analog) which can be modulated appropriately.

An analogy breaks down, then, roughly at the level of specificity at which differences prove to be predominantly of the structure-violating sort. Analogies vary in their *completeness*, that is, their degree of approximation to an isomorphism in which all differences are structure preserving. The usefulness of an analogy, like the usefulness of any mental model, is determined by pragmatic factors. An imperfect analogy can be used to construct rules that provide a first approximation to a valid transition function for the target model. This approximate model will be useful if it can be refined by other inferential mechanisms. The process of refining an approximate solution derived by analogy is conceptually the same as refining a mental model by the addition of new categories and diachronic rules, and can be viewed as an extension of the overall problem-solving task (Carbonell, 1982). In general, the relationship between a source model and a successful target model constructed from it will correspond to a q-morphism, as described earlier.

## D. Critique of the Syntactic Approach

The pragmatic framework predicts that the information transferred from a source to a target will be heavily influenced by the system's goal. The analogist will attempt to construct a set of diachronic rules for the target problem that embodies a transition function adequate to achieve the goal. This characterization of analogy differs sharply from syntactic approaches that attempt to predict the outcome of analogical transfer in terms of purely formal analyses of the structure of the source and target analogs, without making reference to goals. Gentner (1983) is the most emphatic proponent of the syntactic approach. In her theory, "the interpretation rules are characterized purely syntactically. That is, the processing mechanism that selects the initial candidate set of predicates to map attends only to the *structure* of the knowledge representations for the two analogs, and not to the content" (Gentner, 1983, p. 165). A critical examination of Gentner's analysis will illustrate some of the difficulties that generally beset syntactic accounts of inductive mechanisms (Holland *et al.*, 1985).

Gentner distinguishes between "attributes," which are one-place predicates, "first-order relations," which are multiplace predicates with objects as arguments, and "higher-order relations," which are multiplace predicates with propositions as arguments. The syntactic claim is that in using an analogy, people are most likely to map higher-order relations, next most likely to map first-order relations, and least likely to map attributes. For example, in the analogy between atomic structure and a solar system, the target and source share the higher-order relation that "attraction depends on distance," and the first-order relation that "objects revolve around each other." However, attributes of mapped objects, such as their absolute size, do not transfer. Gentner relates the preference for relations, especially higher-order relations, to what she terms a "systematicity" principle. This principle states that a highly interconnected predicate structure—one in which higher-order relations enforce connections among lower-order predicates—is most likely to be mapped.

At first glance Gentner's account looks quite reasonable. Indeed, in the case of interdomain analogies such as that between atoms and solar systems, the primacy of relations in the mapping is virtually definitional. However, a closer examination reveals problems for the syntactic approach. First, as Gentner acknowledges, the higher-order relations of interest typically are such predicates as "causes," "implies," and "depends on," that is, causal elements that are pragmatically important to goal attainment. Thus, the pragmatic approach readily accounts for the phenomena cited as support for Gentner's theory.

Second, it is by no means clear that the systematicity principle distinguishes relations that transfer from those that do not. In the solar system analogy, Gentner mentions the relation "the sun is hotter than the planets" as an example of a relation that fails to transfer because it does not participate in an interconnected set of propositions. This claim is highly suspect. The relative heat of the sun and its planets is causally related to an indefinitely large number of other propositions, such as those describing why only the sun is a star, how the planets originated, the potential for life on the sun versus its planets, and so on. These interconnected propositions obviously have little or nothing to do with our understanding of the analogy with atomic structure, but the systematicity principle is quite unhelpful in showing why they are irrelevant.

Indeed, a basic problem with Gentner's analysis is that it seems to imply that the mappable propositions can be determined by a syntactic analysis of the source analog alone (since the relational status of propositions is defined independently of their participation in an analogy). It follows that the same information should be transferred in all analogies involving a given source. This is clearly false. As an example, let us take as our source analog the concept "elephant." Suppose we know a person of large girth with a penchant for stumbling over furniture. If I were to remark that "Sam is an elephant," the analogical basis of the metaphor would be quite clear. Now suppose that I tell you, "Induction is an elephant." You may be forgiven a moment of incomprehension, especially if you were misled into considering how induction might resemble our clumsy acquaintance Sam. You may grasp my meaning, however, if I remind you of the well-known story of the blind men who grasped different parts of an elephant and then gave totally different descriptions of what the beast is like. Induction may be as much an elephant as Sam, but only the latter has been insulted. Clearly, the basis of an analogy is intimately related not only to the source, but also to the target and the context in which the analogy is used.



Finally, not all analogies are so abstract as that between the solar system and the atom. Consider, for example, an analogy that 4-year-old children can often use to help solve a problem (Holyoak, Junn, & Billman, 1984). The target problem presented to the children required them to find a way to transfer a number of small balls from one bowl to another bowl that was out of reach. A variety of objects were available for possible use, including a rectangular sheet of paper. A source analog, presented prior to the target problem in the guise of an illustrated story, described how the television character Miss Piggy rolled up a carpet and used it as a tube to move her jewels from her jewel box to a safe. Many children were able to generate an analogous solution to the ball problem—rolling the paper to form a tube and then rolling the balls through the tube to the distant bowl.

In this simple analogy, the successful mapping between the carpet and the piece of paper on the attribute of shape is important because in each analog the shape enabled the critical "rolling" operator to be applied. In contrast, the relation of location failed to map (the paper was initially located on the table, whereas Miss Piggy's carpet was initially on the floor). However, this mapping failure was irrelevant because the initial location of the to-be-rolled object was not causally related to the solution achieved in the source analog.

The syntactic approach might attempt to discount this example of attribute precedence in mapping by claiming the story and target problem used by Holyoak *et al.* were too similar to constitute a "real" analogy. However, the resulting restrictive definition of analogy would be quite arbitrary. The pragmatic approach admits of analogies that range over the entire continuum from the mundane to the metaphorical. Even objects that Gentner would term "literally similar" can be analogically related if a goal is apparent. Because Gentner's theory is stated in terms of mappings between static propositions rather than rules, it misses the fundamental distinction between synchronic and diachronic relations, that is, between the mapping and the transition function. An analogy, like any model, must bring with it rules for predicting state changes. Thus, it is anomalous to say that "cats are analogous to dogs" (rather than "cats are *similar* to dogs") because no clear predictive goal is apparent. In contrast, it is perfectly natural to consider whether cats and dogs are analogous *with respect to some specified predictive goal*, such as determining if cats can swim given knowledge that dogs can do so (i.e., predicting the state change that would result from placing a cat in a large body of water).

The syntactic approach is unable to accurately predict the basis for analogical transfer because it fails to take account of goals. Differing goals can lead to different mappings for what is putatively the "same" analogy (Holyoak, 1984b). The perceived structure of an analogy is heavily influenced by the pragmatic context of its use. The aspects of the source analog transferred to the target will be determined by a variety of factors, including knowledge of what aspects of the source are conventionally taken to be important or salient (Ortony, 1979), the apparent goal in using the analogy (e.g., what aspects of the target need to be explained), the causal relations known to be central in the source (Winston, 1980), and what aspects can in fact be mapped without generating structure-violating differences. These complex and interactive factors, which are obscured by purely syntactic analyses of analogy, can be investigated within a pragmatic framework.

## IV. Selection and Use of a Source Analog

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. As noted earlier, the analogy mechanism is most likely to come into play when initial solution efforts based on available diachronic rules that describe the behavior of the target domain fail to generate an acceptable solution plan. In such circumstances it may be necessary to attempt a less direct approach.

In the remainder of this article, I will focus discussion on the processes by which a plausible source analog can be initially selected (the second step of the four specified earlier). The selection step is crucial to the initiation of an explicit mapping procedure. In some cases source analogs are generated by systematic transformations of the target (Clement, 1982). In other cases the source analog will be directly provided by a teacher, as is the case, for example, when the solar system is used as an analogy to elucidate atomic structure. For present purposes, however, the cases of central interest are those in which the analogist notices the relevance of some prior situation to a target problem without external guidance.

The conditions under which people are likely to notice potential analogies are far from clear. Indeed, a consistent research finding has been that college subjects often fail to spontaneously make use of analogies (Gick & Holyoak, 1980, 1983; Hayes & Simon, 1977; Reed, Ernst, & Banerji, 1974). For example, whereas about 75% of the subjects in a typical experiment by Gick and Holyoak were able to generate the convergence solution to the ray problem given a hint to use the prior military story, only about 30% generated this solution prior to receiving an explicit hint. Given that about 10% of sujects produce the convergence solution without any analog, this means that only about 20% of the subjects may have spontaneously noticed and applied the analogy.



In fact, one could reasonably question whether there is any convincing evidence that people notice analogies between problems presented in substantially remote contexts. Even in the case of analogies between problems in the same domain, such as geometry, anecdotal reports suggest that students seldom notice analogies between problems presented in different chapters of their textbook. In all the experiments reported to date, the source and target analogs were presented consecutively within a single experimental session. It could be, for example, that the 20% of subjects in the Gick and Holyoak experiments who spontaneously used the analogy did so simply because they were sensitive to demand characteristics of the situation, which would surely suggest that the story and the problem immediately following might be somehow related.

### A. Schema Induction and Spontaneous Transfer

The strongest evidence that analogs drawn from remote domains are ever spontaneously noticed comes from studies in which multiple source analogs are provided. Gick and Holyoak (1983) had some groups of subjects first read two convergence stories (e.g., the military story described earlier and a fire-fighting story in which converging sources of fire retardant were used to extinguish a large blaze). Other groups read a single convergence story plus a disanalogous story. All subjects summarized each story and also wrote descriptions of how the two stories were similar. The latter task was intended to trigger a mapping between the two stories, which would have the incidental effect of leading to the induction of an explicit representation of the shared schematic structure. All subjects then attempted to solve the ray problem, both before and after a hint to consider the stories.

Gick and Holyoak found that subjects in the two-analog groups were significantly more likely to produce the convergence solution, both before and after the hint, than were subjects in the one-analog groups. Since demand characteristics were presumably comparable for both sets of subjects, the advantage of the two-analog subjects prior to the hint is evidence of spontaneous transfer.

Gick and Holyoak interpreted these and other more detailed results as indicating that induction of an explicit schema facilitates transfer. One might argue, however, that the results simply demonstrate that two analogs are more likely to produce transfer than one because subjects given two analogs have an additional analog that might be retrieved (i.e., if they fail to notice the relevance of one they might nonetheless notice the relevance of the other). However, in a recent experiment, Richard Catrambone and I (1985) have obtained further evidence that schema induction is the basis of the advantage afforded by multiple analogs. We replicated the comparison between groups receiving two analogs versus one analog plus a disanalogous control story. As in the procedure used by Gick and Holyoak (1983), all subjects wrote summaries of each individual story they read. However, we varied whether or not the subjects wrote descriptions of the similarities between the two stories. If this comparison procedure is in fact critical in triggering schema induction (because subjects do not spontaneously perform a detailed mapping between the two stories), then provision of two analogs should only be beneficial when subjects compare the two. On the other hand, if each story is retrieved independently, then two analogs should produce greater transfer than one, even if the stories are not compared to each other.

Table II presents the percentage of subjects in each of the four experimental groups who produced the convergence solution to the ray problem before receiving a hint to use the stories, after the hint, and in total. Twenty subjects served in each group. The results for the comparison conditions closely replicated the comparable conditions run by Gick and Holyoak (1983) in that subjects receiving two rather than one analog were significantly more likely to produce the convergence solution, both before the hint and in total once the hint was given. In particular, 60% of the subjects given two analogs produced the solution spontaneously, versus only 20% of those receiving one analog plus a control story.

The pattern differed for the groups who were not given comparison instructions. In this case only 35% of the subjects given two analogs produced the convergence solution prior to the hint. This figure was significantly lower than

TABLE II

PERCENTAGE OF SUBJECTS PRODUCING CONVERGENCE SOLUTION

| Experimental groups | Before hint | After hint | Total |
|---|---|---|---|
| Two analogs | | | |
| Comparison | 60 | 30 | 90 |
| No comparison | 35 | 45 | 80 |
| One analog plus control | | | |
| Comparison | 20 | 25 | 45 |
| No comparison | 20 | 15 | 35 |



that obtained for subjects in the two-analog comparison condition and did not differ significantly from that obtained for the one-analog conditions. However, once a hint to use the stories was given, the total percentage of subjects producing the solution did not differ as a function of comparison instructions, as both two-analog conditions significantly surpassed the one-analog conditions (overall means of 85 versus 40%).

These results support the schema induction hypothesis regarding the mechanism by which multiple analogs facilitate transfer. At least when the two potential source analogs are superficially dissimilar from each other, as was the case in the above experiment, college students are apparently unlikely to map the analogs and form an explicit schema unless directed to compare them. Consequently, simply providing two analogs, without comparison instructions, does little to foster spontaneous transfer. However, once a hint to use the stories is provided, subjects presumably could retrieve the analogs from memory and map them with each other and the target problem to form a generalized convergence schema. As a result, provision of two analogs aided total transfer once the hint was given, even in the absence of explicit comparison instructions.

It should be noted that in the experiment described above, as in the earlier Gick and Holyoak (1983) study, the groups given one analog in conjunction with a disanalogous control story exhibited lower levels of transfer (especially in terms of total solution frequency) than is typically observed when one analog is given alone. The presence of a "distractor" story makes it more difficult for subjects to identify and apply a source analog. However, the advantage of the two-analog comparison condition in frequency of spontaneous transfer is not attributable solely to the negative impact of the disanalogous story in the one-analog condition. The figure of 60% transfer prior to the hint for the former condition is about twice as great as that consistently obtained in conditions in which one analog is given without a distractor story (Gick & Holyoak, 1980, 1983).

The importance of schemas in mediating transfer across disparate domains can be readily understood in terms of the kind of processing system described earlier. A problem schema is an abstract category that will be linked to a cluster of synchronic rules for categorizing situations of that type and diachronic rules for constructing an appropriate type of solution. Indeed, once a person has induced a schema from initial examples, novel problems that can be categorized as instances of the schema can be solved without directly accessing representations of the initial analogs. It follows that while experiments illustrating the role of schemas demonstrate spontaneous interdomain transfer, they do not provide evidence of *analogical* transfer in the strict sense of direct transfer from a representation of a particular prior situation to a novel problem. Evidence of spontaneous analogical transfer thus remains elusive.

## B. Support Summation and Retrieval of Plausible Analogs

In order to pursue the question of whether and when people spontaneously notice interdomain analogies, let us again consider the operation of the processing system described earlier. When a target problem is established as the focus of processing, the messages describing the problematic situation will direct the ensuing rule-based search for a solution. These messages will describe the initial state, the goal state, relevant operators, and constraints on a solution. Among the rules executed early in the problem-solving attempt will be synchronic rules that support messages describing salient properties of the problem components. Given the ray problem, for example, the message representing "ray" will trigger rules that support messages such as "is a force," "passes through dense matter," "is invisible," and so on. Such property messages will in turn trigger rules supporting messages describing other entities that also share these properties.

The resulting "aura" of associations to the components of the target problem will initially be quite diffuse. However, we assume that multiple sources of support for a message will summate. Whereas a single shared property will likely yield only a small increment in support for a message describing an associated entity, several shared properties will tend to raise its support sufficiently high as to allow it to have an impact on subsequent processing. If messages representing a possible source analog receive sufficient support, rules for constructing a possible mapping with the target will be invoked.

The summation principle tends to ensure that source analogs sharing multiple properties with the target will be activated. It might seem that situations with many superficial similarities will be retrieved. As we will see, surface similarities *do* play a role in analogical retrieval. However, the retrieval process will tend to be dominated by shared properties that are goal related, since the goal in the target problem will in large part determine which rules are executed, and hence which properties of components of the target will actually be activated. Second, plausible source analogs will be those that are related to multiple components of the target problem. In particular, a situation that is activated by both the initial state and the goal state in the target problem is likely to have associated diachronic rules relevant to transforming a corresponding initial state in the source into a corresponding goal state. In other words, a possible analog has been activated when synchronic rules connect the initial target state to an initial state in a source domain, diachronic rules in the source domain connect its initial state to a subsequent state, and the latter is in turn connected by synchronic rules to the target goal. In terms of the quadrilateral diagram in Fig. 3, at this point the system will have begun to develop the mapping between the source and target and to identify diachronic rules that provide the transition function $T'_A$ for the source.



As the above description suggests, the steps of identifying a source analog and performing a mapping between the target and source actually merge. Once a potential source analog is well supported, it will become the focus of continued processing to extend and refine the mapping. This will set the stage for the fourth step in analogical transfer—generation of new rules for the target by substituting mapped elements in the corresponding source rules. If the entire process is successful, these rules will provide a first approximation to an appropriate transition function $T'_B$ for the target.

## C. Surface and Structural Similarities

Given the above description of retrieval of an analog, we can consider why it often seems so difficult for people to spontaneously access relevant source analogs. The basic problem is that a remote analog by definition shares few of the salient features of the target. To the extent the latter serve as retrieval cues, they will tend to activate competing associations that may block retrieval of more remote analogs. The more the problem solver is able to identify and focus on the aspects of the target problem causally relevant to achieving a solution, the greater the probability that a useful but remote analog will be retrieved.

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.

Ideally, a problem solver would use only the structural properties of the target as retrieval cues, thus avoiding activation of superficially similar but unhelpful situations. In reality, however, the problem solver's ability to distinguish structural from surface properties will be at best imperfect, since full knowledge of which properties of the target are structural depends on knowing the possible solutions—information clearly unavailable at the outset of a solution attempt. Consequently, properties that in fact are functionally irrelevant to a solution to the target problem may affect the solution plan indirectly by influencing the selection of a source analog (see Gilovich, 1981).

Once a source analog has been retrieved, surface properties should have less impact on the mapping process than structural ones. In particular, structure-violating differences will necessitate refinement of the initial solution plan generated by the mapping, whereas structure-preserving differences will not. Thus, surface properties will tend to have a relatively greater impact on selection of a source analog than on the subsequent mapping process. For example, it is much easier to learn about atomic structure by mapping it with a solar system than to spontaneously link the two analogs in the first place. In contrast, structure-violating differences will diminish not only the probability of selecting the source analog, but also the probability of using it successfully once mapping is initiated.

Of course, a test of the effects of types of dissimilarity on different steps in the transfer process requires a situation in which subjects in fact sometimes spontaneously notice analogies. As we have seen, there has been little empirical evidence that interdomain analogies are ever spontaneously noticed. However, a recent study by Holyoak and Koh (1985) provides such evidence, setting the stage for investigation of the influence of surface and structural properties on noticing and applying analogies. They investigated transfer between the ray problem and another convergence situation. In the "laser and light bulb" problem, the filament of a light bulb in a physics lab is broken. Because the light bulb is expensive, it would be worthwhile to repair it. A strong laser could be used to fuse the filament; however, it would break the surrounding glass bulb. The convergence solution, of course, is to use several weak lasers focused on the filament.

Relative to the low frequency of spontaneous transfer that Gick and Holyoak (1980, 1983) had found using the military analog described earlier, transfer between the radiation and light bulb problems was excellent. One experiment involved students enrolled in introductory psychology classes. Seventeen experimental subjects were drawn from classes that used a textbook with a detailed discussion of the ray problem, whereas ten control subjects were selected from classes that used texts that did not mention the problem. A few days after the experimental subjects had read about the radiation problem in their textbook as part of a regular assignment, all subjects participated in an experiment (out of class) in which the light bulb problem was presented. About 80% of the subjects who had read about the radiation problem spontaneously generated the convergence solution, as contrasted with a scant 10% of the control subjects who had not. Another experiment revealed that transfer was also good when the light bulb problem was the source and the ray problem the target.

The light bulb analog differs from the military analog described earlier along many dimensions, so it is difficult to determine precisely why the former yields greater transfer. One clear possibility is the difference in the degree of similarity between the instruments of the two analogs and the radiation problem. A laser is obviously far more similar to X rays than an army is, providing a significant additional retrieval cue in the former case. In addition, the deeper structural parallels between the light bulb and radiation analogs make the analogy extremely complete. Both cases involve a target area enclosed within a fragile



"container" that is at risk from a high-intensity force. Thus, both a surface similarity of instruments and a structural similarity of problem constraints provide retrieval cues that can connect the light bulb and radiation analogs.

In an attempt to disentangle the contributions of surface and structural similarities as retrieval cues, Holyoak and Koh generated additional variations of the light bulb analog in which these factors were varied. To vary the surface similarity of the instruments to X rays, two of the new stories substituted "ultrasound waves" for lasers. The problem statement was also altered: Instead of the filament being described as broken apart, it was described as having fused together, and the ultrasound waves could repair it by jarring it apart. Thus, in two stories the solution was to use a laser to fuse the filament, and in two it was to use an ultrasound wave to jar apart the filament. To the extent the two types of action differ in their similarity to that required in the radiation problem (destroying a tumor), the latter appears more similar (since "jarring apart" seems more "destructive" than does "fusing together"). However, the more salient difference is that ultrasound waves are far less associated with X rays than are lasers.[4]

Independently of the variation in the instrument, the stories also varied in their structural similarity to the radiation problem. Specifically, the nature of the constraint preventing direct application of a large force was varied. In the versions with relatively complete mappings, the constraint was similar to that in the radiation problem—a high-intensity force would damage the surrounding area (fragile glass). In the versions with less complete mappings, the constraint was simply that no single instrument of sufficient intensity (laser or ultrasound) was available. These latter versions thus removed a structural cue linking them to the radiation problem. Nonetheless, all four of the stories described essentially identical convergence solutions.

These two types of variations—of instruments and constraints—yielded four alternative stories that were used as source analogs for different groups of subjects. A total of 16 subjects served in each of the two fragile-glass conditions and 15 served in each of the insufficient-intensity conditions. As the data in Table III indicate, the versions differed greatly in their subsequent transfer to the target radiation problem. Table IIIA presents the percentage of subjects in each of the four conditions who generated the convergence solution prior to receiving a hint to consider the story. When the source was the "laser and fragile glass" analog, which has both a similar instrument and a complete mapping, 69% of the subjects spontaneously generated the convergence solution. Transfer was significantly impaired if *either* the surface similarity of the instrument *or* the structural constraint similarity was reduced. If both changes were made (the "ultrasound of

[4]In a further experiment both the laser and the ultrasound beam were described as being able to "jar apart" a fused filament. The pattern of transfer was the same as in the experiment described here.

TABLE III

PERCENTAGE OF SUBJECTS PRODUCING CONVERGENCE SOLUTION

| Structural similarity (constraint) | Surface similarity (instrument) High (laser) | Low (ultrasound) | Mean |
|---|---|---|---|
| | A. Prior to hint | | |
| High (fragile glass) | 69 | 38 | 54 |
| Low (insufficient intensity) | 33 | 13 | 23 |
| Mean | 51 | 26 | |
| | B. Total (before and after hint) | | |
| High (fragile glass) | 75 | 81 | 78 |
| Low (insufficient intensity) | 60 | 47 | 54 |
| Mean | 68 | 64 | |

insufficient intensity'' version), only 13% of the subjects generated the convergence solution. These results indicate that both surface similarities and deeper structural commonalities aid in the retrieval of source analogs, as our earlier account of retrieval mechanisms would predict.

As the data in Table IIIB indicate, a different transfer pattern was observed once a hint to use the story was provided. Structural dissimilarity of the constraint significantly impaired total transfer (78% for the fragile-glass versions vs. 53% for the insufficient-intensity versions), whereas surface dissimilarity of the instruments did not (68% for the laser versions vs. 65% for the ultrasound versions). Thus, although structural and surface similarity had comparable effects on spontaneous transfer, only the former had a significant impact on total analogical transfer once a hint was provided. These results therefore support the prediction that surface similarity will have a greater relative impact on retrieval of a source analog than on application of an analog once it is retrieved.

The results of Holyoak and Koh should not be construed as indicating that surface properties will *never* influence mapping once a source is selected. In the above experiment only a single change was introduced to create a surface dissimilarity. It might well be that introduction of multiple surface dissimilarities would make it more difficult to map the components of the two analogs. In addition, surface differences will continue to impair transfer if the problem solver has difficulty discriminating them from structural differences even after a source analog is provided. In an experiment on analogical transfer performed with 6-year olds, with the ball problem mentioned earlier as the target, Holyoak *et al.* (1984) found that what appeared to be a minor surface dissimilarity between the source and target significantly decreased the percentage of children able to use the analogy. It may be that children, lacking experience with a problem domain, have greater difficulty than adults in analyzing the casually relevant aspects of the source and target problems.



## V. Conclusion

Analogical transfer, I have tried to argue, can be best understood within a broad pragmatic framework for induction. The most fundamental questions regarding analogy concern its roles in a goal-directed processing system. When will analogies be noticed? When can they be put to effective use? What information will be transferred from a source analog to a target problem? How does analogy relate to other problem-solving methods? How does it relate to generalization and other inductive mechanisms? Such questions can be fruitfully addressed only by taking account of the goals of the cognitive system and the principles that govern its capacity to draw inferences and to adapt to its environment.

If this characterization is correct, it implies that it would be a mistake to view the study of analogy as an isolated research area or to anticipate development of a theory addressing analogical transfer alone. An adequate theory of analogy will be forced to make commitments regarding a sweeping range of cognitive mechanisms—attention allocation, knowledge representation, memory retrieval, the dynamics of problem solving, and induction. This conclusion, of course, suggests that understanding analogy will remain an elusive goal for some time. It is certainly the case that the present article falls far short of offering a computationally adequate account of how interdomain analogies can be found and effectively used. More optimistically, however, it is likely that what we learn about analogical transfer will impose important constraints on many aspects of cognitive theories. Research on analogy may prove pragmatically central in our attempts to induce the mechanisms of human cognition.

ACKNOWLEDGMENTS

This article was prepared while the author held an NIMH Research Scientist Development Award, 1-K02-MH00342-04, and the research described was supported by NSF Grant BNS-8216068. Many of the ideas discussed arose in the course of my collaborations with Mary Gick, John Holland, Richard Nisbett, and Paul Thagard (none of whom should be considered responsible for the vagaries of my exposition). I am grateful for the facilities provided by the Psychology Department of Carnegie-Mellon University, where I was visiting when the article was written. The article benefited from comments on an earlier version provided by Miriam Bassok, Matthew Lewis, Peter Pirolli, and Jeff Shrager, as well as from discussion in John Anderson's research seminar.

## References


Anderson, J. R., Greeno, J. G., Kline, P. J., & Neves, D. M. (1981). Acquisition of problem-solving skill. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition.* Hillsdale, NJ: Erlbaum.

Carbonell, J. G. (1982). Learning by analogy: Formulating and generalizing plans from past experience. In R. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach.* Palo Alto, CA: Tioga Press.

Catrambone, R., & Holyoak, K. J. (1985). *The function of schemas in analogical problem solving.* Poster presented at the meeting of the American Psychological Association, Los Angeles, California.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology,* in press.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1985). *Pragmatic versus syntactic approaches to training deductive reasoning.* (in preparation).

Clement, J. (1982). *Spontaneous analogies in problem solving: The progressive construction of mental models.* Paper presented at the meeting of the American Educational Research Association, New York.

Duncker, K. (1945). On problem solving. *Psychological Monographs,* **58** (Whole No. 270).

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

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.

Gilovich, T. (1981). Seeing the past in the present: The effect of associations to familiar events on judgments and decisions. *Journal of Personality and Social Psychology,* **40,** 797–808.

Hayes, J. R., & Simon, H. A. (1977). Psychological differences among problem isomorphs. In N. J. Castellan, Jr., D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory.* Hillsdale, NJ: Erlbaum.

Hesse, M. B. (1966). *Models and analogies in science.* Notre Dame, Ind.: Notre Dame Univ. Press.

Holland, J. H. (1975). *Adaptation in natural and artificial systems.* Ann Arbor, Mich.: Univ. of Michigan Press.

Holland, J. H. (in press). Escaping brittleness: The possibilities of general purpose learning algorithms applied to parallel rule-based systems. In R. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 2). Palo Alto, CA: Tioga Press.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1985). *Induction: Processes of inference, learning, and discovery* (in preparation).

Holyoak, K. J. (1982). An analogical framework for literary interpretation. *Poetics,* **11,** 105–126.

Holyoak, K. J. (1984a). Mental models in problem solving. In J. R. Anderson & S. M. Kosslyn (Eds.), *Tutorials in learning and memory: Essays in honor of Gordon Bower.* San Francisco: Freeman.

Holyoak, K. J. (1984b). Analogical thinking and human intelligence. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2). Hillsdale, NJ: Erlbaum.

Holyoak, K. J., Junn, E. N., & Billman, D. O. (1984). Development of analogical problem-solving skill. *Developmental Psychology,* **55,** 2042–2055.

Holyoak, K. J., & Koh, K. (1985). *Surface and structural similarity in analogical transfer* (in preparation).

Newell, A., & Simon, H. A. (1972). *Human problem solving.* New York: Prentice-Hall.





Ortony, A. (1979). Beyond literal similarity. *Psychological Review,* **87,** 161–180.

Pirolli, P. L., & Anderson, J. R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology,* in press.

Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology,* **6,** 436–450.

Reitman, W. (1964). Heuristic decision procedures, open constraints, and the structure of ill-defined problems. In M. W. Shelley & G. L. Bryan (Eds.), *Human judgments and optimality.* New York: Wiley.

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