

Learning Physical Domains:

Toward a Theoretical Framework

Kenneth D. Forbus & Dedre Gentner
University of Illinois at Urbana-Champaign

Abstract

This chapter presents a theoretical framework that is being developed in an attempt to construct a computational account of human learning of physical domains. Qualitative Process theory is used to model portions of people's physical knowledge, and Structural Mapping theory is used to characterize the computations that move a learner from one representation to another. The chapter outlines the component theories and proposes a learning sequence for physical domains.

12.1 INTRODUCTION

People use and extend their knowledge of the physical world constantly. Understanding how this fluency is achieved would be an important milestone in understanding human learning and intelligence, as well as a useful guide for constructing machines that learn. The authors' purpose is to construct a computational account of human experiential learning in physical domains.

This work is still at the stage where questions are being refined rather than answers provided. In many cases, there is no direct evidence for the claims made here. In other instances, support for the theory is obtained by combining evidence from several different areas, including developmental psychology, studies of learning, and other psychological research. No one of these is adequate by itself. When extrapolating from adult learning research, we must keep in mind that cases of pure experiential learning are rare in adult life; some sort of instruction or prior expectation is typically involved. Developmental research provides a good source of data, since much of young children's learning is truly from direct experience.

Yet when developmental results are applied, it must be remembered that children are not only learning but also maturing. Therefore, in order to isolate and study experiential learning, the existing empirical findings must be examined, filtered, and carefully fitted together. Although space does not permit detailing all the relevant lines of evidence, the authors will try to give the reader some justification for our claims whenever possible.

As this volume attests, the past few years have seen significant progress in machine learning. However, to construct programs that learn as well as (or better than) people do, it is important to understand how human learning works. Ultimately both psychological studies and direct computational experiments (i.e., constructing programs) will be necessary to provide a full account. To this end, the authors will try when possible to indicate how techniques developed in machine learning might be used to implement such programs.

12.1.1 Overview

A brief prologue may help to organize the material. Three key ideas underlie the theory: (1) the centrality of *physical processes* in mental models of science; (2) the importance of analogy in learning; and (3) the primacy of rich, contextually specific representations. The idea that the notion of process is central to human knowledge about physical domains is the chief tenet of Qualitative Process (QP) theory (Forbus, 1981, 1984). This is not to say that notions of process are there from the beginning. Rather; it is hypothesized that a person's experiential knowledge of a domain begins as a collection of scenarios that describe particular phenomena, out of which is developed a vocabulary of processes that provides a notion of mechanism for the domain. The second key idea concerns the role of comparisons among related knowledge structures. The authors conjecture that much of experiential learning proceeds through spontaneous comparisons-which may be implicit or explicit-between a current scenario and prior similar or analogous scenarios that the learner has stored in memory. Structure Mapping theory (see Gentner, 1980, 1983) describes these kinds of comparisons.

The third idea is a rather paradoxical claim: in human processing, more is often easier.¹ Rich, perceptually based representations are acquired earlier in learning than sparse abstract representations. That is, early domain representations differ from more advanced representations of the same domain in containing more information, especially perceptual information specific to the initial context of use.

¹It should be noted that psychologists by no means generally agree with this claim. Consequently, the authors tried to be fairly explicit in presenting evidence for this position.

A second aspect of the "more is easier" claim concerns comparisons: it is suggested that, for humans, similarity comparisons are easier when there is more overlap between the two knowledge structures being compared.

On the basis of these three ideas, the authors propose a canonical learning sequence. The claim is that human learning of physical domains can be viewed as a sequence of different mental models: (1) *protohistories*, (2) the *causal corpus*, (3) *naive physics*, and (4) *expert models*. Briefly, protohistories are rich, contextually specific, highly perceptual representations of phenomena, capturing expectations about typical phenomenological patterns—for example, "If I turn the key, the car will start." With the causal corpus, the expectations of mechanism enters; here the representation consists of simple statements that some sort of causal connection exists between variables—"If the car has no gas, it will not start." In the naive physics stage, processes are introduced to provide the mechanism underlying the causal corpus—"Gas must flow from the tank to the carburetor and mix with air so that the mixture can be ignited by the spark." The disparate local connections of the causal corpus are replaced with qualitative models organized around the notion of process. Finally, in the expert models stage, quantitative representations are created—for example, models of the effects of different mixtures of oxygen and gasoline.

In this chapter the authors discuss their conjectures about these models and about how a learner constructs one type of model from another. First, however, the component theories that underlie this framework are briefly summarized: Qualitative Process theory, which provides concepts needed to represent the models (particularly in the naive physics stage), and Structure Mapping theory, which characterizes the kinds of computations that move the learner from one representation to another. Then the overall role of structure-mapping comparisons is examined in the progression from rich to sparse representations. With these foundations in place, the four stages of learning for physical domains are then described.

12.2 QUALITATIVE PROCESS THEORY

The first requirement of this work is a language in which to describe people's commonsense knowledge about physical situations. People know about a great many kinds of physical changes: things move, collide, bend, break, heat up, cool down, flow, and boil. Intuitively we think of these as *processes*. Qualitative Process theory attempts to formalize this notion of process to provide a common form for qualitative theories of dynamics. As will become clear later on, the authors do not believe that the first models people construct of a domain take the

form of process, nor even that they become knowledgeable enough to construct these models for every domain they experience. Nevertheless, some of the concepts of QP theory will be useful in describing models in other stages as well.

In QP theory, a physical situation is modeled as a collection of objects and relationships among them, with processes responsible for causing changes. The continuous parameters of an object, such as temperature and pressure, are represented by *quantities*. A quantity has two parts, an *amount* and a *derivative*. Amounts and derivatives are both *numbers*. The model to keep in mind for numbers is that of the reals, but it is important to note that in QP theory particular numerical values are never used. Instead, the value of a number is described in terms of its *quantity space*-a collection of inequalities that hold between it and other quantities. Figure 12-1 illustrates a quantity space for the level of liquid in a container. The quantity space is a useful qualitative representation because processes typically start and stop when inequalities between parameters change.

Figure 12-2 illustrates a typical process, called LIQUID-FLOW. A process has five parts: *individuals*, *preconditions*, *quantity conditions*, *relations*, and *influences*. Roughly speaking, the individuals part describes where instances of a process might occur, the preconditions and quantity conditions tell when it will be acting, and the relations and influences describe what holds as a consequence of it acting. In more detail, for any collection of objects that matches the individual specifications there is a *process instance* that represents the potential for that process to occur between those individuals in a particular way. For example, there will be two instances of LIQUID-FLOW between the liquid in the containers in figure 12-2, each corresponding to flow in a particular direction.

A process instance is *active* whenever both its preconditions and its quantity conditions are true. The distinction between preconditions and quantity conditions is that quantity conditions can be determined within QP theory but preconditions cannot. Quantity conditions concern what inequalities hold and what other processes (or individual views, which are introduced below) are active. Preconditions concern any relevant factors other than quantity conditions, such as spatial boundaries. For example, in "real" physics we can solve equations to figure out how fast a ball will be moving when it hits the floor, but the equations will not tell us a priori where the floor is. Or, returning to the present example, if we know that all the valves in the fluid path between the two containers are open (i.e., the fluid path is aligned), then fluid will flow, but we cannot predict within QP theory when or if someone will walk by and turn off a valve. Because these factors still affect dynamical conclusions, preconditions must be explicitly represented.

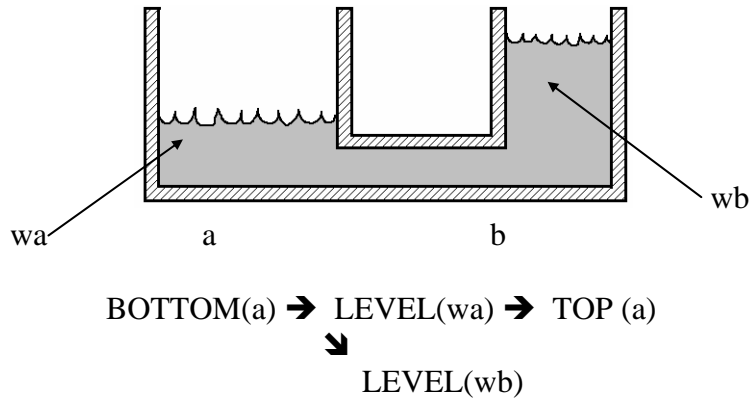


Figure 12-1: A quantity space describes the value of a number by the inequality relationships that hold between it and other numbers. An arrow indicates that the number at its head is greater than the number at its tail. Thus $\text{LEVEL}(wa)$ is less than $\text{LEVEL}(wb)$ and greater than $\text{BOTTOM}(a)$, and $\text{LEVEL}(wb)$ and $\text{TOP}(a)$ are unordered.

Process LIQUID-FLOW

Individuals:

source, a CONTAINED-LIQUID
 dest, a CONTAINED-LIQUID
 path, a FLUID-PATH, FLUID-
 CONNECTION(source, dest, path)

Preconditions:

ALIGNED(path)

Quantity Conditions:

$A[\text{PRESSURE}(\text{source})] > A[\text{PRESSURE}(\text{dest})]$

Relations:

Let flow-rate, diff be quantities
 $\text{diff} = \text{PRESSURE}(\text{source}) - \text{PRESSURE}(\text{dest})$
 $\text{flow-rate} \propto_{Q+} \text{diff}$

Influences:

$I + (\text{AMOUNT-OF}(\text{dest}), A[\text{flow-rate}])$ $I -$
 $(\text{AMOUNT-OF}(\text{source}), A[\text{flow-rate}])$

Figure 12-2: A typical process. This process specification describes a simple kind of liquid flow. It can occur between two contained liquids that are connected by a fluid path, whenever the path is aligned—that is, all valves in the path are open—and the pressure in the one taken as source is greater than the pressure in the contained liquid taken as destination. The quantity type AMOUNT-OF represents how much "stuff" there is in an object. (Recall the function of A maps a quantity into the number that is its amount, a number, as opposed to AMOUNT-OF, which is a function that maps a piece of stuff into a quantity.)

Whenever a process instance is active, its influences and relations hold.

The influences component of a process specifies the direct effects of a process; the relations component describes other facts that are true while the process is active. The direct effects-called *direct influences*-take the form

$$I + (Q, n)$$

or

$$I - (Q, n)$$

depending on whether n is a positive or negative contribution to the derivative of Q . If a quantity is directly influenced, its derivative will be the sum of all the direct influences on it. Returning to the description of LIQUID-FLOW, for example, we see that when an instance of LIQUID-FLOW is active, there will be a positive influence on the amount of liquid in the destination and an equal, negative influence on the amount of liquid in the source.

The relations field can describe new individuals that are created by virtue of the process being active (such as the steam produced by boiling water), as well as properties needed by representations outside QP theory (such as the appearance of boiling water). An especially important kind of fact expressed in the relations component is functional dependency between quantities. Functional dependencies between quantities are expressed by

$$Q1 \propto_{Q+} Q2$$

(read "Q1 is *qualitatively proportional* to Q2," or informally, "Q1 q-prop Q2"), meaning there exists a function that determines Q1 and is strictly increasing in its dependence on Q2. \propto_{Q-} indicates that the dependence is strictly decreasing. Note that qualitative proportionalities express partial information, since the exact nature of the function relating the parameters is not known and the function may or may not depend upon other quantities.² If a quantity Q1 is functionally dependent on a quantity Q2, and if Q2 is influenced by a process P, then we will say that P *indirectly influences* Q1; that is, when P is acting it can cause Q1 to change. If, for instance, the PRESSURE and LEVEL of a liquid are qualitatively proportional to the AMOUNT-OF of the liquid, then LIQUID-FLOW will indirectly influence both PRESSURE and LEVEL because it directly influences AMOUNT-OF. It is important to note that the only way a quantity can change is if it is directly or indirectly influenced. This means that one can reason by exclusion: If nothing is

²QP theory also provides ways to specify dependence on properties that are not quantities (such as shape, in relating the level of a liquid in a container to its volume) and to make stronger statements about functional relationships, such as "Q1 depends on Q2 directly, with no intervening parameters" and "Q1 depends on Q1 and Q2 and nothing else" when required for framing stronger hypotheses about a domain. However, precise specifications of functions (e.g., $Q1 = |Q2|^2$) are not permitted.

influencing the amount of fluid in a container, then it isn't changing, but if the amount is changing, something must be influencing it. No changes happen by themselves. Furthermore, we can trace the possible paths of influences in a situation and determine whether or not particular kinds of changes can occur.

Two other important types of descriptions should also be mentioned here. *individual views* are descriptions used to represent both objects whose existence are subject to dynamical constraints and states of objects. "The water in a cup," for example, is described as a CONTAINED-LIQUID (see figure 12-3) because we can get rid of it by

INDIVIDUAL-VIEW CONTAINED-LIQUID

Individuals:

c a CONTAINER
s a SUBSTANCE

Preconditions:

CAN-CONTAIN-SUBSTANCE (c, s)

Quantity Conditions:

A[AMQUNT-OF-IN (s, c)] > ZERO

Relations:

THERE IS g, a PIECE-OF-STUFF
HAS-QUANTITY (g, AMOUNT-OF)
AMOUNT-OF (g) = AMOUNT-OF-IN (s, c)
HAS-QUANTITY (g, LEVEL)
LEVEL (g) α_{Q+} AMOUNT-OF (g)
HAS-QUANTITY (g, PRESSURE)
PRESSURE (g) α_{Q+} LEVEL (g)

Figure 12-3: This typical individual view describes a piece of liquid in a container, using the ontology for liquids described in Hayes (1979a). THERE IS is just "syntactic sugar" for stating that whenever the preconditions and quantity conditions are true, g will exist.

reducing its amount to zero (perhaps by making it the source of an instance of LIQUID-FLOW). Another example is a model of a spring. Springs have three states-relaxed, compressed, and stretched-each of which can be modeled by individual views. Individual views are specified in the same way as processes, in that they have individuals, preconditions, quantity conditions, and relations.

However, they do not have an influence component; directly influencing quantities is the sole prerogative of processes.

The other kind of description is the *encapsulated history*. How an object changes through time is represented by its *history* (Hayes, 1979b). Histories are annotated pieces of space-time; thus they are object centered, have finite spatial extent, and extend over time.³ As its name suggests, an encapsulated history is a schematized description of some fragments of histories for a collection of objects. Encapsulated histories are useful for summarizing behavior and for directly describing phenomena that have not been accounted for by process descriptions. An example of the latter usage is describing collisions between moving objects. A very simple way to model collisions is to say that the very next thing that happens after, say, an object hits a wall is that its velocity reverses and it starts moving the other way. Given how rapidly collisions occur, this model is quite adequate for most purposes, and encapsulated histories allow it to be written this way.

A reasoner's theory of dynamics for a particular domain characterized in terms of (1) a *process vocabulary* that describes the kinds of processes the reasoner believes can occur and (2) a *view vocabulary* that describes dynamical objects and relevant states of objects. All changes are assumed to be directly or indirectly caused by processes—the *sole mechanism* assumption—which provides a strong constraint on the form of dynamical theories. Importantly, the content of dynamical theories is not tightly constrained—incorrect theories can be expressed as easily (and sometimes more easily!) than correct theories. For example versions of Newtonian, Aristotelian, and Impetus theories of motion have all been encoded using QP theory.

QP theory sanctions several basic deductions. For example, the kinds of processes that might occur in a situation can be determined by using the process and view vocabularies to construct instances representing the different possibilities. The collection of processes acting at any time characterizes "what is happening" then in that situation, and these processes can be found by evaluating the preconditions and quantity conditions for these instances.

Consider again the example in figure 12-1. There will be two instances of the LIQUID-FLOW process, and since the level in w_b is greater than W_a , the LIQUID-FLOW instance representing flow from w_b to w_a will be active. By taking into account all the influences on each quantity (called *resolving* its influences), we can often determine the sign of its derivative. The sign of the derivative is important because it represents how the amount of the quantity is

³By contrast, the classic situational calculus description of change (McCarthy and Hayes, 1969) consists of situations that describe the whole universe at some particular instant of time.

changing- increasing, decreasing, or remaining constant. In this example there is only one process instance acting, which makes things simple. AMOUNT-OF(wb) is directly influenced, and since this influence is negative, it will decrease. By the α_{Q+} statements in the CONTAINED-LIQUID description, LEVEL(wb) and PRESSURE(wb) will be indirectly influenced and thus will also decrease. Similarly, AMOUNT-OF(wa), LEVEL(wa), and PRESSURE(wa) will increase.

From the ways the quantities are changing we can determine how the process and view structures themselves might change, since they depend in part on the inequalities stated as quantity conditions. This computation is called *limit analysis*. In the example two things might happen- the pressures in wb and wa might equalize, or AMOUNT-OF(wb) could become zero, thus ending wb's existence (the geometry of this example rules out the latter).

The basic deductions of QP theory can be combined to perform more complex reasoning tasks. Two examples of more complex deductions are qualitative simulation (Forbus, 1984) and measurement interpretation (Forbus, 1983). Qualitative simulation consists of performing limit analysis repeatedly. It is useful for making predictions, for instance, that boiling water in a sealed container could cause an explosion. Measurement interpretation provides a link between physical theories and observations; for example, it might be hypothesized that the level of a fluid in a container is dropping because the fluid is flowing out somewhere. Measurements taken at a single instant may be interpreted by searching through the space of process and view structures for situations in which the results of influence resolution match the observations. Algorithms for interpreting measurements taken over a span of time are still under development.

12.3 COMPARISONS AND STRUCTURE MAPPING

So far this chapter has considered how portions of a person's knowledge about the physical world might be represented. Let us now turn to the question of how such domain models might be learned. The authors conjecture that a major process in experiential learning is comparing the current situation with stored descriptions. Consider the example of a person who has just moved to a cold climate and is learning to operate a furnace. Suppose that at first he wrongly believes that the house will get warm faster if the thermostat is set to a temperature higher than the desired temperature. (Kempton shows that this view is quite common; Kempton, 1985.) How can he reach the correct conclusion that the *rate* of heating does not depend on the temperature setting? There are at least three different ways, each based on a different kind of implicit comparison. First, he could compare his past furnace experiences with each other and notice a regularity in the rate of heating

that is independent of the thermostat setting. Second, he may compare the furnace situation with known abstractions and realize that it is best described as a positional-action controller (as opposed to a proportional-action controller). Third, he may use an analogy, comparing the furnace situation with a description from another domain, such as fluid flow, to suggest governing principles. Each of these ways of learning relies on some form of comparison, either with a stored record of literally similar events, with a stored abstraction, or with a stored description that can function as an analogy.

Structure Mapping theory is concerned with such comparisons (see Gentner 1980, 1982, 1983; Gentner and Gentner, 1983). The theory describes the rules that are used to import a descriptive structure from one domain (the *base* domain) into another (the *target* domain). The central intuition is that an analogy implies that a predicate structure from one domain can be applied in another domain with arbitrarily different objects and surface appearances. *Literal similarity*, *analogy*, *mere appearance mappings*, and *abstraction mappings* (applications of general laws) are viewed as different kinds of mappings between descriptions. The types of comparisons are defined syntactically, in terms of the form of the knowledge representation, not in terms of its content. Each type of comparison will be considered in turn.

1. An analogy is a comparison in which relational predicates, but few or no object attributes, are mapped from base to target. The particular relations mapped are determined by *systematicity*, as defined by the existence of higher-order constraining relations that can themselves be mapped.⁴ The correspondences between objects of the base and objects of the target are thus determined by the roles of the objects in the relational structure, not by any intrinsic similarity between the objects themselves.
2. A literal similarity statement is a comparison in which a large number of predicates, both attributes and relations, can be mapped from base to target. Here, the model is based on one proposed by Tversky (1977), which states that the similarity between A and B increases with the size of the intersection of their feature sets and decreases with the size of the intersection of the two

⁴Object attributes are predicates that take one object as an argument, such as RED(x) . Relations are predicates that take two or more arguments, such as COLLIDE (x, y). We define the order of a proposition as follows: Constants have order zero, as do functions on them. The order of a proposition is one plus the maximum of the orders of its arguments. Thus COLLIDE(x, y) would be first order if x and y are domain objects, and CAUSE(COLLIDE(x, y), BREAK(x)) would be second order. Examples of higher-order relations are CAUSE and IMPLIES.

complement sets.⁵ Thus, there are many more shared predicates than nonshared predicates.

3. An abstract mapping is a comparison in which the base domain is an abstract relational structure. Predicates from the abstract base domain are mapped into the target domain. As in analogy, the mapped predicates are a relational structure. Abstraction differs from analogy in the nature of the base domain. There are almost no object attributes in the base, so there are few, if any, one-place predicates to be left behind. Applying a rule to a situation is an example of abstraction mapping. Sometimes the relational structure so mapped will also be referred to as an *abstraction*.
4. A mere appearance match is a comparison in which the object attributes match but the relational structure does not. In a sense it is the opposite of analogy. Such matches are easily made, but they guarantee nothing beyond similarity in appearance.

A series of related examples using the analogy between heat flow and water flow will illustrate these distinctions. Figures 12-4a and 12-4b show a water-flow situation and the corresponding heat-flow situation (adapted from Buckley, 1979, 15-25). Figure 12-5 shows a possible representation a person might have of the water situation. Notice that the description contains both object-attribute predicates, such as CYLINDRICAL(beaker), and relational predicates, such as GREATER-THAN

⁵Again according to Tversky, the negative effects of the two complement sets are not equal; for example, given the question How similar is A to B?, the set (B - A)-features of B not shared by A-counts more than the set (A - B).

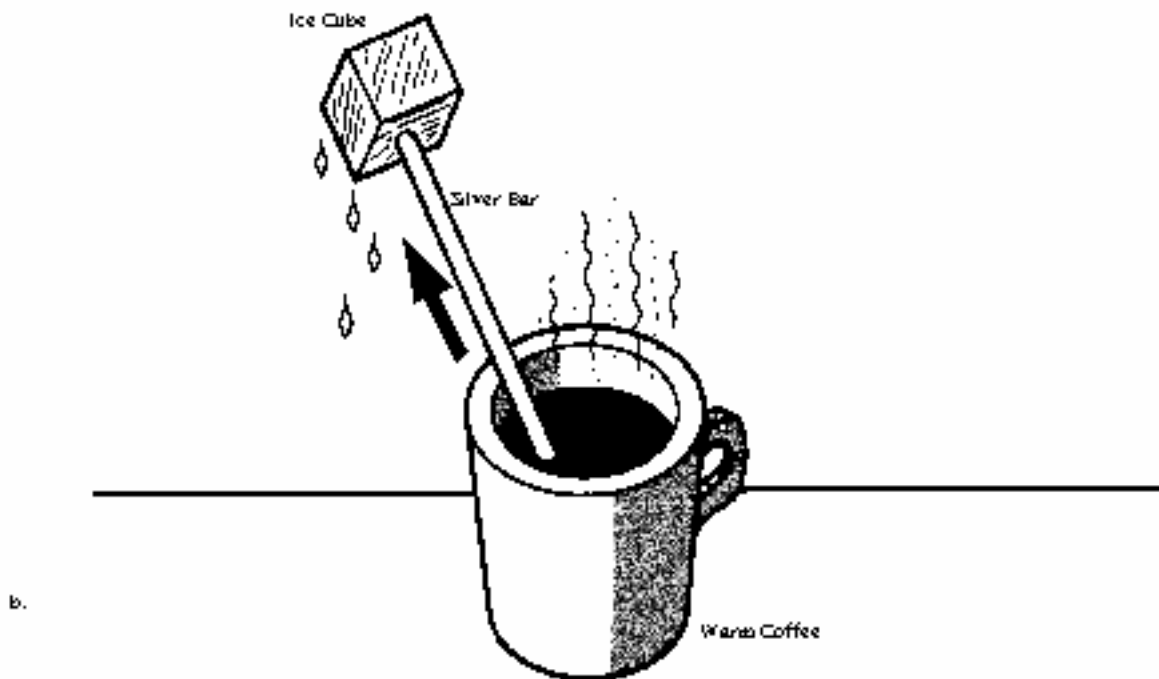
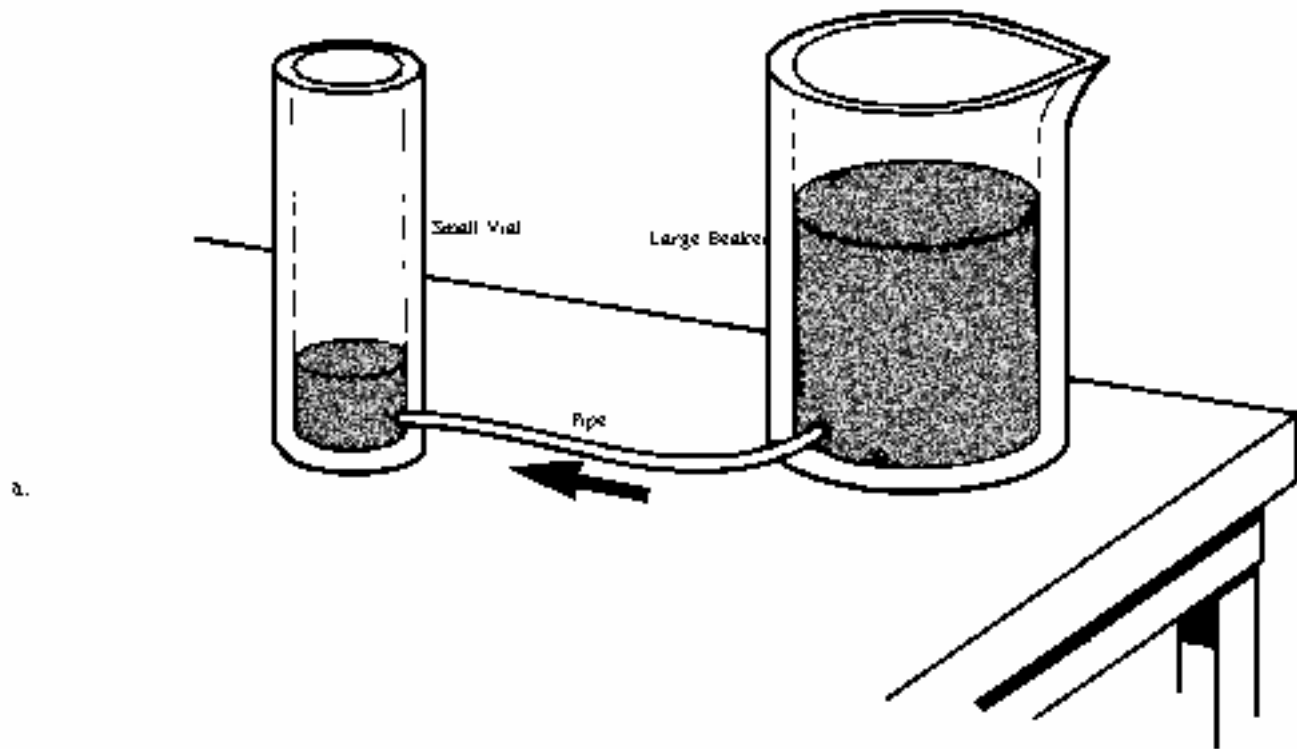


Figure 12-4: These two physical situations involving flow will be used to illustrate the kinds of comparisons sanctioned by Structure Mapping theory and later to illustrate how QP-style domain descriptions can be used in analogies: (a) A water-flow situation; (b) the corresponding heat-flow situation.

[PRESSURE(water, beaker), PRESSURE(water, vial)]. Let us consider the comparison types as exemplified here:

1 The analogy *Heat is like water* conveys that certain aspects of the water description can be mapped onto the heat domain. In particular, (1) object attributes should be dropped; (2) some relational predicates should be carried over; and (3) systematicity determines which relations should be mapped. Thus, CYLINDRICAL(beaker) is dropped, along with other object attributes;

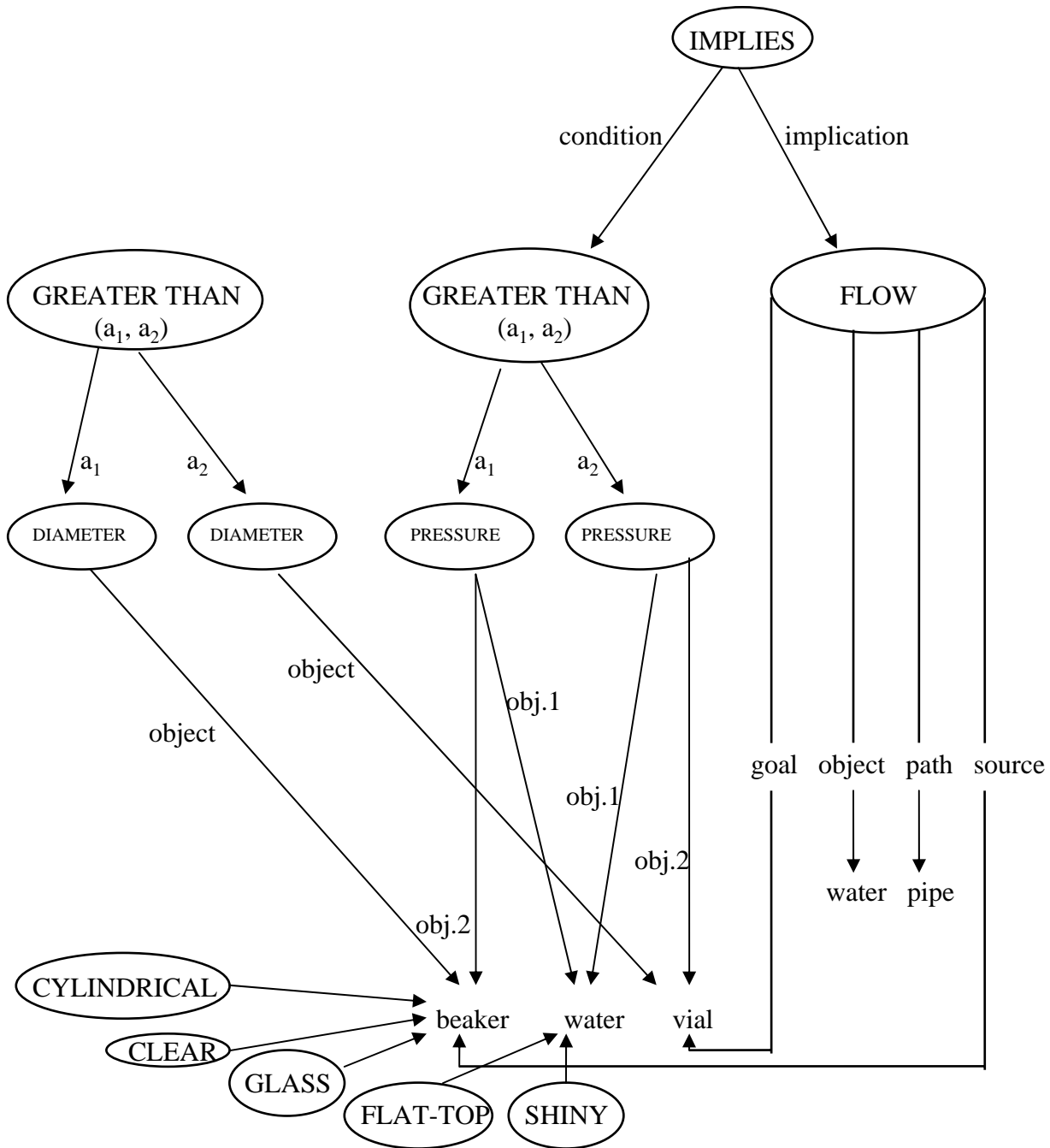


Figure 12-5: A representation of the water situation. This network represents a portion of what a person might know about the water situation illustrated in figure 12-14. In this and other figures, predicates are written in upper case and circled. Objects are written in lower case and uncircled. A simplified representation is used to illustrate the rules of analogy. A more detailed model will be shown.

that is, the target objects do not have to resemble their corresponding base objects. Some relations are carried across, such as, GREATER-THAN [PRESSURE(water, beaker), PRESSURE(water, vial)]. Yet not all relations are carried across. By the systematicity principle, this GREATER-THAN relation is preserved because it is part of the mappable chain governed by higher order relation IMPLIES. In contrast, the relation GREATER-THAN[DIAMETER (beaker), DIAMETER(vial)] is not carried across, since it is not part of any mappable system of constraining relations in this representation of the base domain.

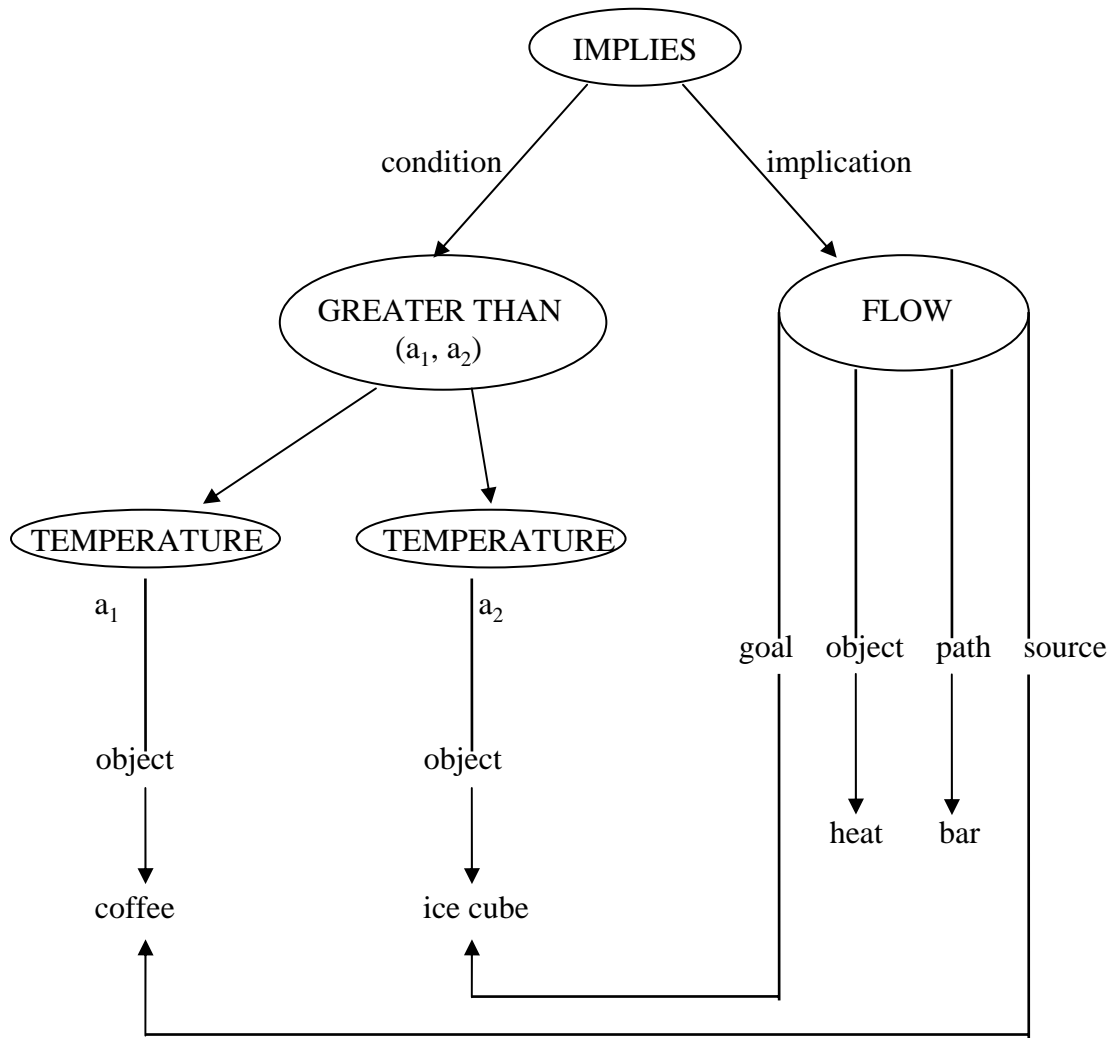


Figure 12-6: A representation of the heat situation that results from the heat/water analogy. This network represents the knowledge a person would map across into the heat domain from the water situation illustrated in figure 12-5. As in that figure, a simplified representation is used here. A more detailed treatment of this analogy is presented later.

Figure 12-6 shows the representation in the target domain of heat flow that results from the analogical mapping. Given the arbitrary object correspondences heat/water, beaker/coffee, vial/ice cube, pipe/bar, and PRESSURE/TEMPERATURE,⁶ systematicity operates to enforce a tacit preference for coherence and predictive power. The systematic relational structure in the water domain

```
IMPLIES (GREATER-THAN [PRESSURE (water, beaker),
                        PRESSURE (water, vial)],
        FLOW (water, pipe, beaker, vial))
```

is mapped into

```
IMPLIES (GREATER-THAN [TEMPERATURE (heat, coffee),
                        TEMPERATURE (heat, ice cube)],
        FLOW (heat, bar, coffee, ice cube))
```

2. The literal similarity comparison *Kool-Aid is like water* conveys that most of the water description can be applied to Kool-Aid. In literal similarity both object attributes, such as FLAT-TOP (water), and relational predicates, such as the systematic chain discussed above, are mapped over.
3. The abstraction *Heat is a through-variable* might be available to a student who knows some system dynamics. This abstraction conveys the idea that heat can be thought of as something that flows across a difference in potential (i.e., some sort of across-variable-in this case, temperature). This is much the relational structure as conveyed by the analogy in 1, above; the difference is that in the abstract base domain of through-variables and across-variables there are no concrete properties of objects to be left behind in the mapping.
4. A mere appearance match is a match with overlap chiefly in lower-order predicates, such as object attributes, but little or no relational match. An

⁶In this analogy, the first-order predicate of PRESSURE in the water domain must be replaced by TEMPERATURE in the heat domain. Although systems of relations can often be imported into the target without change, substitutions of lower-order relations, as well as of objects and their attributes, are sometimes made in order to permit mapping a larger systematic chain.

example is *The tabletop gleamed like water*. Such a match typically yields little or no useful information about the target; here, for example, little can be learned about the table by mapping across knowledge about water. These matches, however, cannot be ignored in a theory of learning, because a novice learner may be unable to tell them from true literal similarity matches.

Table 12-1 summarizes the kinds of predicate overlap that characterize literal similarity, analogy, abstraction, and mere appearance matches, as well as one other kind of comparison, *anomaly*. An anomaly is a match with little or no predicate overlap; it is included simply for completeness.

It should be clear that the contrasts described here are continua, not dichotomies. For example, analogy and literal similarity lie on a continuum. Given that two domains overlap in relational structure, then the comparison becomes more a literal similarity match to the extent that their object attributes also overlap, and more an analogy to the extent that few or no object attributes overlap. A different sort of continuum exists between analogies and general laws. In both cases, a relational structure is mapped from base to target. If the base representation included concrete

OBJECT	ATTRIBUTES	RELATIONS	EXAMPLE
Literal Similarity	Many	Many	Milk is like water.
Analogy	Few	Many	Heat is like water.
Abstraction	Few	Many	Heat flow is a through-variable.
Anomaly	Few	Few	Coffee is like the solar system.
Mere Appearance	Many	Few	The glass tabletop gleamed like a pool of water.

objects whose individual attributes must be left behind in the mapping, the

comparison is an analogy. As the object nodes of the base domain become more abstract and variablelike, the comparison is seen more as a general law.

12.4 STRUCTURE MAPPING AND LEARNING

The role of a comparison in learning depends on at least two things: (1) *accessibility*-the likelihood that the match will be noticed-and (2) *usefulness*- what can be deduced from the match if it is accessed. Accessibility, in turn, depends at least on (a) the *familiarity* of the base description and (b) the overall *similarity* between the base description and the current target. The immediate usefulness of a match depends, of course, on whether the content of the match is appropriate to the task at hand. In addition, the usefulness of a match depends on the *inspectability* of the matching content- the degree to which it can be consciously analyzed and articulated. The comparisons discussed above behave very differently with respect to accessibility and inspectability.

For novice learners, literal similarity matches are the most accessible comparisons, and abstractions are the least accessible. In contrast, abstraction matches are far more inspectable than literal similarity matches. On both dimensions, analogies are intermediate. That is one reason that analogy is crucial in learning. Some evidence for these conjectures will now be reviewed.

Literal similarity matches are highly accessible. It has been shown in education and training literature that the more similar a new situation is to an original situation the more readily transfer of training occurs (cf. Brown and Campione, 1985). The term *generalization gradient* expresses the fact that a learned response generalizes more readily the more similar the new situation is to the original situation. In contrast, subjects are often quite slow to use an available analogy. In research done by Reed, Ernst, and Banerji (1974) and later by Gick and Holyoak (1980, 1983), subjects were asked to solve a rather difficult problem, such as how to cure an inoperable tumor with radiation without killing the flesh along the path of the rays. Just prior to receiving the problem some of the subjects read material that contained an analogous solution, such as a story about a general who split his troops up so that they all converged simultaneously on a fortress he wished to capture

There are three interesting results here. First, a good analogy can be very powerful if it is noticed. Without the analogy, only about 10 percent of the subjects could solve the problem. Once the experimental subjects were told to use the prior story as an analogy, 80 to 90 percent of them solved the problem correctly. Second, a potentially powerful analogy can easily go unnoticed. Before the analogy was

pointed out, only about a third of the subjects spontaneously noticed and used it. It cannot be taken for granted that a potential analogue will be spontaneously noticed and used. Third, literal similarity is far more accessible than true analogy. In one of their studies, Gick and Holyoak (1983) happened to set up a literal similarity match between the story and problem. Subjects had to solve a problem that involved tying two ropes together, and the story they were given involved tying two ribbons together. In this case, 70 to 80 percent of the subjects were able to access the matching story spontaneously.

There is also developmental evidence that literal similarity matches appear prior to analogies and abstraction matches in learning. One example is early word learning. In spontaneous labeling, one-year-old children frequently apply words to objects that closely resemble the original referent of the word; for example *doggie* will be applied to another dog or to a cat, and *car* to cars, trucks, or other vehicles (Clark, 1973). Truly analogous or metaphorical usages are seldom heard until the age of two or three years; for example, a three-year-old child remarks about his dirty bedraggled blanket, "It's out of gas" (Gentner and Stuart, 1984; Winner, 1979).

Children are said to move from rich, concrete representations to more abstract, rule-based systems (cf. Bruner, Olver, and Greenfield, 1966; Gibson, 1969). Even three-year-olds can sort objects into perceptually similar categories; for example, they can group a cat and a dog and exclude a hen. However, not until they are five or six years old can they succeed if the match is more abstract; for example, a category like "living thing" requires grouping perceptually dissimilar things.

In the same vein, research on the novice-expert shift in adult learning has demonstrated that whereas novice science students typically match situations on the basis of surface features, experts use deeper and more abstract criteria (Larkin, 1983). For example, Chi, Feltovich, and Glaser (1981) have shown that when novice physics students are asked to classify problems into similar groups they put together problems with similar surface features, such as "inclined planes" or "pulleys." Experts, on the other hand, use categories like "force problems" and "energy problems."

One final indication of the ease with which literal similarity matches are made involves an indirect, but very important, line of argument. In the realm of object concepts, there is some evidence that people automatically perform literal similarity comparisons to combine perceptually similar experiences into composite prototypes (see Posner and Mitchell, 1967; Rosch, 1973, 1975, 1978; Smith and Medin, 1981).⁷ In the Posner and Mitchell study, people classified dot patterns

⁷The term *prototype* has been used in various ways in psychology. Here it is used to refer to a *structured*

into categories. After they had sorted the patterns, they were asked to remember which patterns they had seen. Although the task called simply for accessing verbatim memory, subjects showed systematic misrecognitions: they falsely remembered having seen prototypical patterns that were never presented. Thus without being told to do so, people formed composite mental representations, apparently based on implicit comparisons among the patterns that they saw. The virtually automatic nature of prototype learning is further evidence that the literal similarity matches on which they are based are highly accessible—indeed, evidence that making such comparisons is a passive, essentially automatic process (see also Reber, 1967, 1976).

However, prototypes also illustrate the limited usefulness of literal similarity matches. Although these implicit composites are often sufficient for recognizing and categorizing situations, they are of limited use in deriving causal principles. This is because (1) a match based largely on perceptual commonalities will often fail to contain the correct principles and (2) even when some of the correct relations are present, literal similarity matches are too rich to be inspectable. There is some evidence, albeit indirect, for this notion of rich, noninspectable representations. Nickerson and Adams (1979) studied people's memory of the common penny. Despite the overwhelming amount of experience that the subjects had with pennies, and despite their evident ability to recognize and categorize pennies, they were remarkably poor at recalling or recognizing the details of how pennies look. This demonstrates that possessing a description sufficient to recognize a class of objects is no guarantee that the description can be articulated.

Studies of young children show that similarity judgments can be difficult to decompose. Shepp (1978) has found that three- and four-year-olds appear to base their similarity judgments on some kind of overall comparison; they are typically unable to judge one dimension independently of another. For example, they cannot ignore height when judging width. Unlike adults, they are unable to treat length and width as separable.

By contrast an appropriate abstraction match is likely to be extremely useful in both respects: it should contain the correct principle, and the match should be inspectable. But abstractions are often not particularly accessible, especially for novices. Novice learners may not know the appropriate abstraction, or it may be so unfamiliar that they will not retrieve it when appropriate. Thus abstraction mappings, although ultimately important, are unlikely to play a major role in the early stages of learning.

Analogies lie between the highly accessible literal similarity matches and the

highly useful abstraction matches. Potential analogies are less accessible in experiential learning than literal similarity matches. This is because analogy requires that the learner's database be accessed via relational matches; object matches are of little or no use. However, once found, an analogy should be more useful than a literal similarity match in deriving the key principles, since the shared data structure is sparse enough to permit analysis. (Of course, educators often explicitly introduce analogies in teaching beginners for exactly this reason, In this case, the problem of noticing the analogical match is bypassed.) Moreover, by the systematicity principle, the set of overlapping predicates is likely to include higher-order relations, such as causality and logical implication. Thus analogy can function to reveal principles in a domain that previously lacked the appropriate abstractions (Burstein, chap. 13 of this volume; Carbonell, chap. 14; Clements, 1982; Darden, 1983; Gentner, 1980, 1982, 1983; Gentner and Gentner, 1983; Gick and Holyoak, 1983; Hoffman, 1980). Winston's system (see Winston, 1980, 1982), which derives if-then rules by abstracting the predicates common to two analogues, is a case in point.

The *analogical shift hypothesis* concerns the role of these comparison in experiential learning. In the earliest stages most of the spontaneous matches are either mere appearance matches (and thus erroneous) or literal similarity matches based on massive feature overlap. This is to say that initial learning is surface oriented and conservative, based on rich, specific-case kinds of matches. As the domain becomes familiar, more distant comparisons begin to occur; matches are made in which fewer object attributes are shared. These sparse comparisons lead to the kinds of binary connections that form the bulk of the causal corpus- for example "Lighter farther when thrown." Analogy also serves as a means of introducing structured mental models. Successful analogies may yield abstractions that can be stored and accessed (Gick and Holyoak, 1983; Winston, 1980, 1982). Thus, analogy plays an important role in the middle and later stages of learning. In the final stages, when learning is well advanced, abstraction mappings play a major role.

12.5 STAGES OF UNDERSTANDING

The authors suspect that four kinds of mental models are generated in the process of understanding physical domains. The sequence of models proposed here is developmental, in that the theories of each stage are generated both by the phenomena being understood and by the theories of the stage before it. It is not proposed that every person go through every stage for every domain, nor that a

person is at the same stage in every domain at the same time.

12.5.1 Stage 1: Protohistories

Suppose some new physical phenomenon is being observed. If there is no prior model, all one can do is observe and remember what is happening. The authors conjecture that the simplest physical models of a domain are *protohistories- prototype* histories that serve as summaries of experience.⁸ Like object prototypes, protohistories are the "most typical instances" of phenomena. The terms in these descriptions are observables, and their deductive import can be roughly expressed as, "If I see X, then Y will happen (has happened)."

Consider a balance beam or seesaw. If a weight is placed on each side of the fulcrum, the seesaw will either tilt counterclockwise, tilt clockwise, or not tilt at all. Most people have had enough experiences with seesaws to have formed protohistories concerning their behavior. By the conjecture described here, a protohistory is automatically available whenever they encounter a seesaw. From it, they can often predict which way the particular seesaw will move. For example, they may have a protohistory that describes what happens if a small person gets on the seesaw opposite a large person.

However, the predictive power of protohistories is quite limited. There is no guarantee that the features matched actually correspond to relevant factors. For example, an observer will be fooled when a large person sits close to the fulcrum if the observer's seesaw protohistories have been formed from watching people sitting at equal distances. Massive overlap in features is needed for reliable use, which means protohistories will yield conclusions in fewer situations than a true theory would. Consider, for example, two weights hung from opposite ends of a stick that is suspended by a string. The principle involved is the same, yet the situations look dissimilar enough that the protohistories for seesaws would not match. Furthermore, there is no certain way to decide between conflicting results if more than one protohistory matches a situation.⁹

12.5.1.1 Learning Protohistories

⁸Some of diSessa's *phenomenological primitives* (1983) appear to be representable as protohistories.

⁹There are of course heuristic criteria, such as using the protohistory that has worked most often. The problem with such heuristics is that little is learned from mistakes.

The process of constructing protohistories involves dividing up experience into classes according to literal similarity and abstracting a summary for each class. There has been little direct research on this process. However, investigations into the process of constructing object prototypes provide some hints. First, people seem to be able implicitly (i.e., unconsciously) to compute a kind of component match. Second, this intersection is not merely a simple feature intersection; rather, it appears that configurations among features are important in the prototype. Third, once this prototype is computed, it has powerful effects on the subsequent processing of experience. As mentioned previously, once people abstract a prototype from a set of patterns, they may be more confident of having seen the prototype—which was never presented—than they are of having seen the patterns actually presented (Posner and Mitchell, 1967). Finally, people may not be aware of forming prototypes, except as a general sense of increased familiarity with a category.

In summary, if protohistories behave like object prototypes, then they should be found to (1) be computed implicitly; (2) act as composite concepts; (3) be sensitive to perceptual configurations among events; and (4), once computed, show the recognition strength and other psychological privileges of prototypes.

The machine learning research that most closely captures this type of learning is concerned with *conceptual clustering* (see Michalski and Stepp, 1983). So far, such research has focused on classifying objects that can be characterized mainly by differing attributes. Extending such techniques to describe situations that depend critically on relational descriptions could provide a method for computing protohistories (Stepp and Michalski, chap. 17).

12.5.2 Stage 2: The Causal Corpus

Protohistories summarize the phenomena, but they do not constitute a theory of them. Building a detailed theory directly can be quite difficult. The space of possible models connecting all observable (and possible) parameters in a typical situation can be quite large. The authors conjecture that weaker theories, theories that characterize which parts of the situation are relevant to desired conclusions, are formed first. In particular, it is conjectured that a collection of CAUSE statements, the causal corpus, is computed from prototype objects and protohistories.

CAUSE is viewed here as an approximate concept, a weak form of ontological commitment. In particular, saying CAUSE(A, B) expresses belief in the existence of some mechanism, specified by some theory T, such that IMPLIES

[(AND A T), B]. Many, perhaps most, of the causal corpus relations are binary relations among variables—for example, "Bigger objects weigh more" (Piaget, 1951; Carey, 1985), or "Smaller objects have higher pitch when struck" (diSessa, 1983).

The notion of mechanism in the causal corpus is quite primitive: the causal beliefs need be neither explicit nor internally consistent. Later in the learning sequence, as will be seen, *processes* will assume the role of mechanisms for physical domains. Nevertheless, the authors conjecture that even at this early stage the learner makes a distinction between mechanistic connections and, say, definitional connections.¹⁰ Further, they suspect that many of the initial causal connections are incorrect. Novices often include diagnostic and correlational relations in their causal corpus. For example, asked if an increase in the evaporation rate will cause a change in the temperature of the water, a novice may reply, "Yes, because it would have to be hotter to evaporate more."

CAUSE, then, is a statement of belief in some mechanistic connection. The distillation of experience from protohistories into the causal corpus serves three purposes. First, it serves as a means of data reduction. Second, it provides a collection of heuristics that can be used directly to draw inferences. Even if the learner doesn't have firm grounds to consider the CAUSE statements complete or correct, the CAUSE statements may often suffice for the desired class of inferences. Third, the collection of CAUSE relations can be used to guide the search for a deeper theory of the domain. The CAUSE statements suggest connections among various aspects of the domain that a deeper theory must either explain or explain away.

Returning to the seesaw example, suppose the causal corpus is now applied to a balance beam built out of blocks. Suppose the two blocks on it are called a and b. The causal corpus might be as follows:

CAUSE (BIGGER (a, b), TILT-TOWARDS (a))
CAUSE (FARTHER (a, b), TILT-TOWARDS (a))

These statements can be interpreted as rules in several ways: If block a is bigger than block b, one can predict tilt, and if one sees tilt, one may hypothesize that one block is farther out than another. These statements are more broadly applicable

¹⁰For example, the statement CAUSE(TRIANGLE(f), HAS-THREE-SIDES(f)) is not a legitimate use of CAUSE by this account, since the required axioms of geometry do not specify a mechanism.

than protohistories since they refer to fewer properties. Unlike protohistories, the causal corpus is sparse enough to be debugged to some degree.

However, the approximate nature of the CAUSE relation limits the learner's ability to discriminate between conflicting predictions. With the causal corpus above, for instance, if block a is bigger and block b is farther out, we will have two predictions. Inhelder and Piaget (1985) and Siegler (1976, 1981) have documented such a stage in the development of understanding about the balance beam (with analogous developmental sequences in other domains). Initially, children focus only on weight. But there is an interesting second stage when they come to realize that both weight and distance are important but they do not yet know the interrelations. They can manage either property by itself if the other is constant; but if both properties vary, they tend to focus on one or the other inconsistently. Eventually they become able to coordinate weight and distance in the balance beam problem. At this stage, if not before, they have gone beyond the causal corpus. As will be discussed, in order to make more precise inferences the learner must eventually uncover the mechanisms whose behavior is described by causal corpus.

12.5.2.1 Learning the Causal Corpus

The authors suspect that there are three techniques for computing and debugging a causal corpus. The first technique is to hypothesize causality from co-occurrence:

If you always see A before B,
then hypothesize CAUSE (A, B)

and

If A is true whenever B is true,
then hypothesize CAUSE (A, B)

These rules make certain assumptions about the form of memory, namely, that some number of circumstances can be remembered and that they can be remembered in sufficient detail that A and B are either explicitly stored or computable from what is stored. Protohistories should serve as a means of initial data reduction from which a causal corpus can be constructed.

It is not clear exactly how the learner abstracts out particular variables from

the rich representation of the protohistory stage. However this is done, the simplification achieved with the causal corpus is considerable. Another study by Siegler (1978) shows the power of focusing on particular variables. Three-year-old children were shown a balance beam, asked to predict which way it would tilt, and then shown what actually occurred. Even after large numbers of trials, their performance failed to improve. But when they were taught to think of the domain in terms of a few relevant variables-weight and length-their performance did improve with experience. The moral to be drawn is that the pace of learning is greatly accelerated when a small number of variables can be abstracted from all the possibly relevant factors.

As suggested earlier, many of the early causal relations will be incorrect. The authors suspect that there exists a class of rules that are used to debug a causal corpus in the face of new information (cf. Sussman, 1976). Each rule corresponds to a hypothesis about a bug in the structure of the causal corpus, such as a missing precondition. The authors believe that the task of judging a causal corpus for consistency is an example of an important but relatively neglected kind of learning, *coherence-driven learning*. Coherence-driven learning is learning that is driven not by a mismatch between the model and the world but by inconsistencies within the model itself. Williams, Hollan, and Stevens (1983) found evidence of such learning. They studied a subject who was learning about a heat exchanger and noted that one source of insight was a "boggle" experience, in which the person noticed that a current inference contradicted a prior belief. The authors are still examining the criteria for judging the consistency of a causal corpus.¹¹ Such criteria will play a major role in controlling the debugging rules and the mixture of generation and debugging that occurs.

Analogy provides the third technique for extending a causal corpus (see Gentner and Gentner, 1983; Stevens, Collins, and Goldin, 1979). The CAUSE relations from one domain can be mapped into another, since CAUSE qualifies as a higher-order constraining relation (see also Winston, 1982).

12.5.3 Stage 3: Naive Physics

The naive physics models replace CAUSE statements with theories about the specific mechanisms of change. The ontology is extended by adding processes to explain observed changes. The ontology also includes properties and objects that

¹¹With Lance Rips of the University of Chicago, the authors are investigating the role of intransitives in debugging causal descriptions.

are not directly observable (for example, heat and heat flow) and the new relationships (such as fluid path and heat path) required to reason about them.

An important advantage of these models is that they allow one to reason by exclusion. Unlike the previous stages, predictions that fail still yield information about the situation. For instance, if fluid is flowing into a container and the level is not rising, then it is reasonable to hypothesize that fluid is flowing out of it through some unknown path.

Returning to the balance beam example, a process SWING might be used to describe rotation around a contact point (see figure 12-7). The preconditions describe the geometric configuration of the system, and the quantity condition says that SWING will occur whenever there is a nonzero angular velocity. SWING directly influences the angular position of the beam. Thus a prediction concerning tilt becomes a prediction about which instance, if any, of the SWING process will be active.¹²

What influences ANGULAR-VELOCITY? The existence of an ANGULAR-ACCELERATION process (see figure 12-8) that directly influences ANGULAR-VELOCITY whenever there is a net torque will be assumed. It is further assumed that

```
For-All (x) For-All (y)
  PHYSOB(x) and CONTACT-POINT(cp)
  implies NET-TORQUE(x, op) = SUM-OF(TORQUES-
ON(x, cp))
```

```
Process SWING
Individuals:
  b a PHYSOB
  c a PHYSOB
  op a CONTACT-POINT
  dir a DIRECTION
```

```
Preconditions:
  MOBILE(b)
  not MOBILE(c)
```

¹²An alternate, and equally good, representation for SWING would leave directions implicit in the sign of the velocity. In that vocabulary, the balance beam would give rise to only one instance of SWING, and determining which way the beam moves requires that one determine first whether the instance of swing is active and if it is, what the sign of the angular velocity is.

```

CONNECTED(b, C, op)
ROTATION-FREE(b, C, op)
DIRECTION-OF(dir, ANOULAR-VELOCITY(b, op))

```

Quantity Conditions:

```
An[ANGULAR-VELOCITY(b, op)] > ZERO
```

Influences:

```
I + (ANGULAR-POSITION(b, op), A[ANGULAR-
VELOCITY(b, op)])
```

Figure 12-7: A SWING process describes rotation of an object around another object. For the balance beam there will be two instances of this process, differing only in their bindings for the direction dir. In each instance b will be bound to the beam, c will be bound to the fulcrum, and op will be bound to the contact point between them.

It is assumed that each physical object (PHYSOB) has quantities to represent its angular position and velocity with respect to each point of contact with other objects. Directions will be noted by the symbols CW, CCW, and NULL, corresponding to clockwise rotation, counterclockwise rotation, and no rotation.

In other words, the net torque on an object around a contact point is the sum of the torques on that object measured around that contact point. The mass of the beam will be ignored, and the pull of gravity on the blocks on each side of the fulcrum will be assumed to be the only source of torques. Figure 12-9 describes this induced torque by means of an individual view. Notice that the factors illuminated in the causal corpus of BIGGER and FARTHER have become the quantities MASS and DISTANCE, and their role in producing swinging has been explicated. In particular, these properties determine how much torque each block places on the beam. The sum of the torques determines the net torque, which can cause the beam to accelerate and thus swing.

This model comes one step closer to a model that can always determine which way something will tilt. There will still be cases in which exactly what will happen cannot be determined (e.g., if the mass on one side is increased and it is brought closer to the pivot), but this is a precise hypothesis about what all the relevant factors are.

12.5.3.1 Learning Naive Physics

The major problem in learning a naive physics is constructing a vocabulary of processes that consistently describes experience. The learner must strip away the

Process ANGULAR-ACCELERATION

```

Individuals:
b a PHYSOB
c a PHYSOB
cp a CONTACT-POINT
dir a DIRECTION

Preconditions:
MOBILE(b)
not MOBILE(c)
CONNECTED(b, c, cp)
ROTATION-FREE(b, c, cp)
DIRECTION-OF(dir, NET-TORQUE(b, cp))

Quantity Conditions:
Am[NET-TORQUE(b, op)] > ZERO

Relations:
Let acc be a quantity
acc  $\propto$  Q + NET-TORQUE(b, op)
acc  $\propto$  Q- MASS(b)

Influences:
I + (ANGULAR-VELOCITY(b, op), A[acc])

```

Figure 12-8: An ANGULAR-ACCELERATION process.

irrelevant predicates that are part of his or her protohistories and causal corpus and construct more appropriate descriptions. In addition, the learner must sometimes hypothesize the existence of objects and properties that are not directly observable. Research in machine learning has developed several techniques for inductive learning that should prove useful (see Dietterich and Michalski, 1983; Mitchell, 1982 Michalski, 1983). These problems are starting to be addressed directly in the study of scientific discovery (Langley, et al., 1983).

The causal corpus provides a search space for potential process vocabularies. Each statement in the causal corpus must be elaborated into a consequence of a process vocabulary. It appears that there are only a small number of distinct ways

to perform the elaboration, depending on the particular form of the arguments. For example, the statement

The decrease in AMOUNT-OF q
causes the LEVEL OF Q to fall

Individual View GRAVITY-INDUCED-TORQUE

Individuals:

b a PHYSOB
 c a PHYSOB
 d a PHYSOB
 cp a CONTACT-POINT

Preconditions:

CONNECTED(b , c , cp)
ON(d , b)

Relations:

Let f be a quantity
 f ELEMENT-OF TORQUES-ON(b , cp)
 $f \propto Q + \text{DISTANCE}(C-M(d), op)$
 $f \propto Q \text{ MASS}(d)$
;Assign positive torques to CW, negative
torques to CCW
ON(C-M(d), SIDE-OF(CW, b , cp)) iff As [f] = 1
ON(C-M(d), SIDE-OF(CCV, b , cp)) iff As [f] = -
1
ON(C-M(d), SIDE-OF(NULL, b , cp)) iff As [f] =
0

Figure 12-9: A description of gravity-induced torque.

indicates that some active process (or individual view) in the situation contains the statement

$\text{LEVEL}(q) \propto Q + \text{AMOUNT-OF}(q)$

in its relations.

Hypothesizing a process vocabulary from a causal corpus should be much simpler than working from protohistories or direct observation. Yet it still appears difficult. The authors conjecture that there are several constraints that make the problem more tractable. First, people are apparently *conservative* in the introduction of unobserved properties. For example, some subjects have a model of a domain that appears to be organized around one parameter—a "generalized strength" attribute. In reasoning about fluids, for instance, they appear to use *pressure*, *flow rate*, and *velocity* as different names for the same thing. In electricity, they use *voltage*, *current*, *power*, *potential*, and *velocity of electrons* interchangeably. The advantage of this theory generation strategy is, of course, that simpler models will be explored first, with further distinctions made only when necessary. Second, some physical laws are used as constraints on what process vocabularies are possible. Conservation of energy, for example, demands that if a process directly influences a quantity representing some form of energy, it must also directly influence some other quantity representing some form of energy, but in the opposite direction.

Once again, analogy can provide a constructive mechanism. It can be used to import candidate processes from previously understood domains—for example, when one understands heat flow in terms of fluid flow. This is an especially powerful mechanism because if the model for the previous domain is consistent with physical laws, then it suggests that the model for the new domain may be so as well. Recall the liquid flow model presented in section 12-2. Figure 12-10 illustrates a collection of assertions that describes the consequences of a particular instance of LIQUID-FLOW.¹³

Suppose a person hypothesizes that there is a process of heat flow analogous to the process of liquid flow. By Structure Mapping theory, this means that the person suspects that a similar relational structure holds among the objects in the heat-flow situation (the coffee, the ice cube, the silver bar, and the instance of heat flow) as holds among the objects in the liquid-flow situation (the water in the beaker, the water in the vial, the pipe, and the instance of liquid flow). Mapping the systematic relational structure (see figure 12-11) leads to several predictions that the person can check to see whether the analogy is correct. For example, it can be determined whether or not the temperature of the ice cube is rising and the temperature of the coffee falling. The structure-mapping rules for analogy have

¹³The assertions were generated by an early version of GIZMO, a computer program being constructed to *explore* the computational aspects of QP theory. GIZMO was designed to make predictions and interpret measurements, not to be a learning system. In particular, these descriptions were not generated with learning or analogy in mind.

provided an initial model for the process of heat flow; in particular, the preconditions, quantity conditions, relations, and influences are all carried across from liquid flow. Note that to make the analogy really work, a new kind of object—a HEAT-PATH—must be postulated. Thus analogy can provide candidates for extending ontologies.¹⁴

12.5.4 Stage 4: Expert Models

The models generated so far have two important limitations. First, they still contain fundamental ambiguities, ambiguities that are inherent in the nature of qualitative representations.¹⁵ Second, they lack domain-independent generalizations (except in the raw form of the representation-CAUSE statements, processes, and so on). The final stage of learning consists of overcoming these limitations, of discovering ways to resolve ambiguities, and to construct powerful generalizations.

Clearly several kinds of knowledge are involved, and the potential complexity of the models in this stage is open-ended (it includes the whole of mathematical physics, for example). Examples of the kinds of knowledge involved include

¹⁴Of course, such extensions are not to be made lightly. The authors suspect that new types of objects are Postulated in the target domain only when necessary to preserve a much larger systematic structure.

¹⁵The nature of ambiguity in qualitative descriptions is discussed by deKleer (1979) and Forbus (1984).

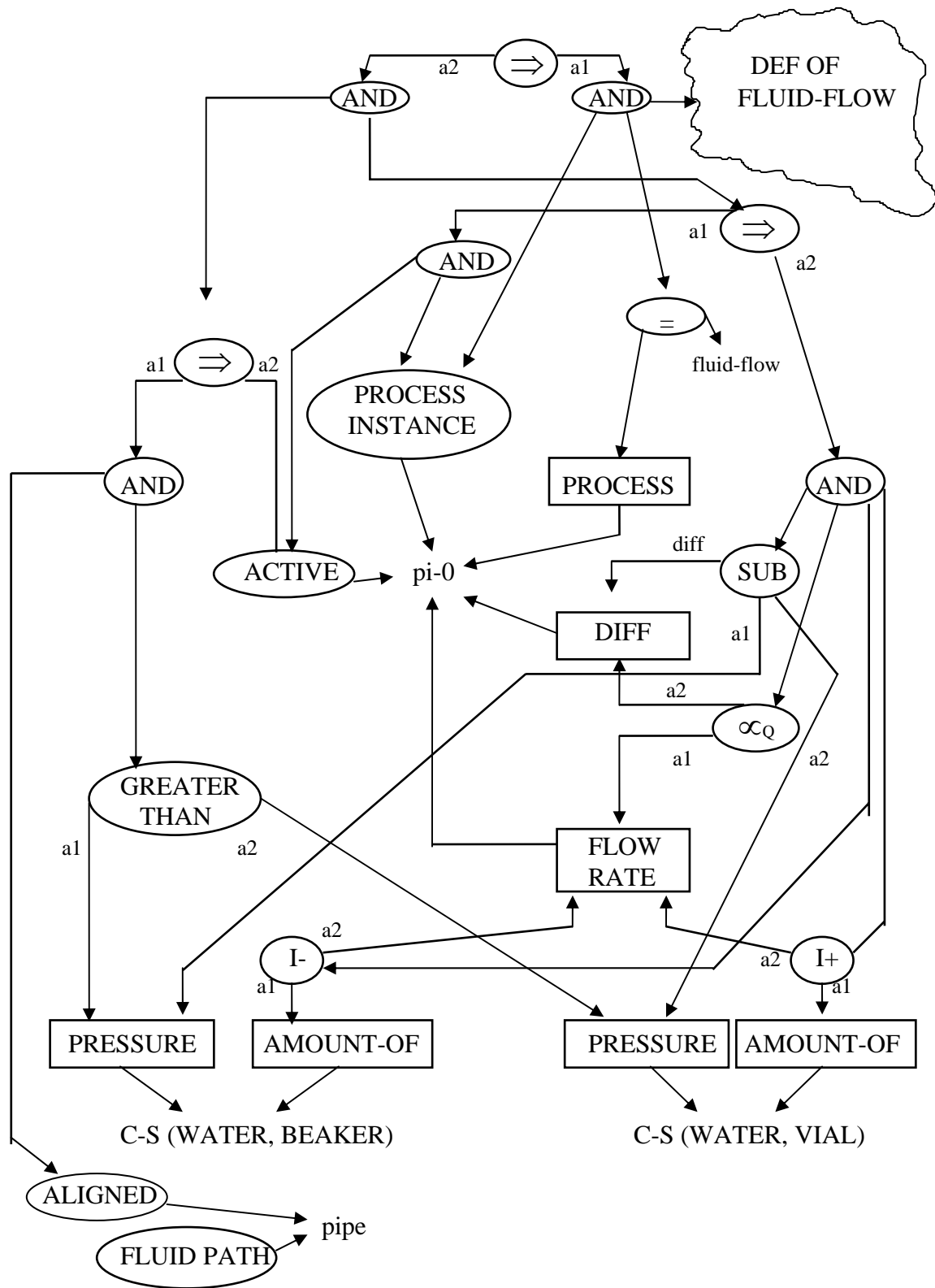


Figure 12-10: Relational structure for an instance of liquid flow. These important conclusions follow from the definition of liquid flow presented in figure 12-2 and the assumption that an instance of liquid flow exists involving the liquids in the two containers. Specifically, they describe the conditions for and consequences of the process instance $pi = 0$'s being active.

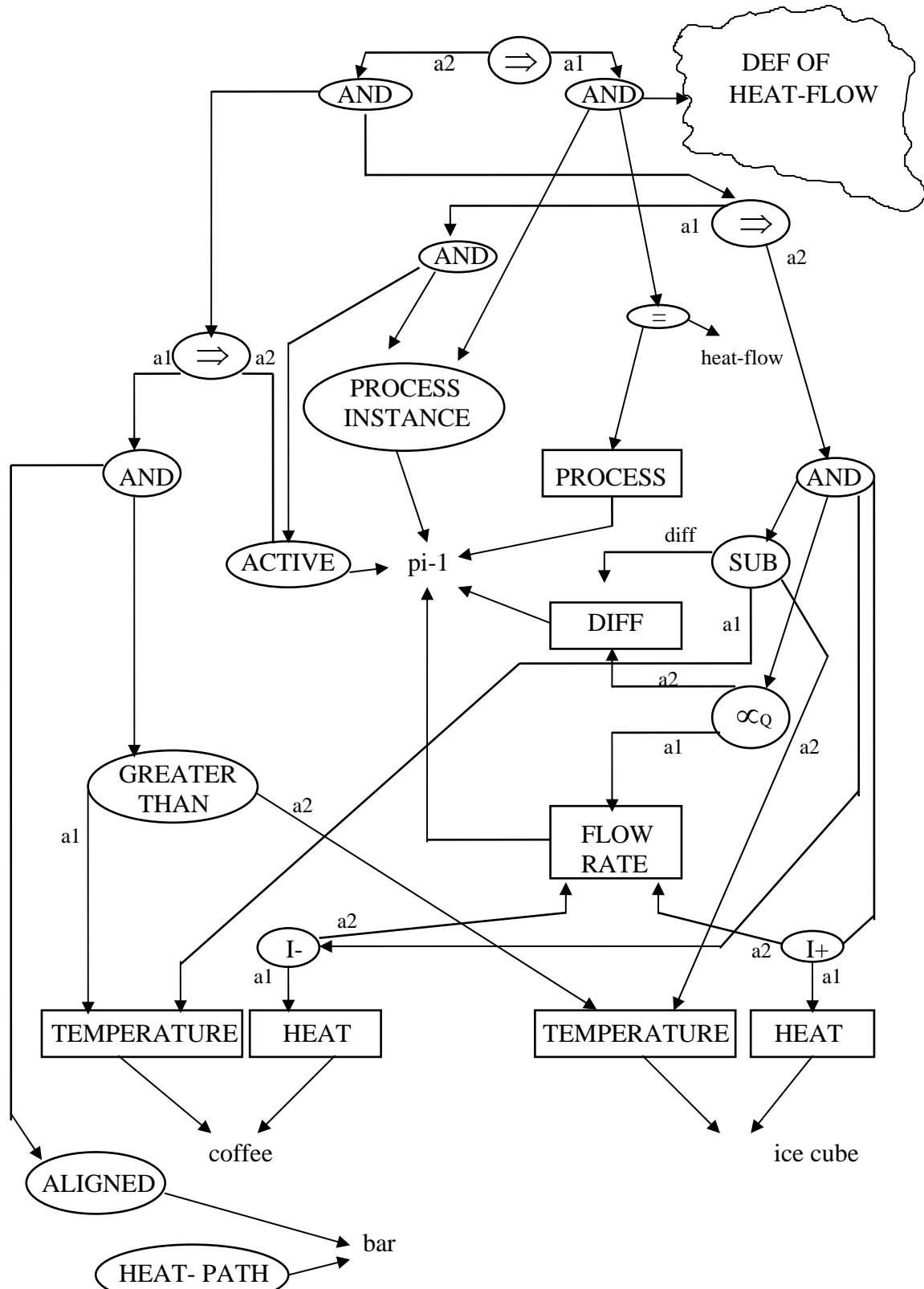


Figure 12-11: Relational structure transferred to heat flow. Here the relational structure describing a situation involving liquid flow has been transferred to a situation involving heat flow. Notice the systematicity of the relational structure, as indicated by the nested chains of implications.

equations to describe relationships between parameters, "rules of thumb" to specify useful default resolutions for ambiguities, and new ontologies to allow reasoning about more complex systems. The importance of mathematical models is fairly obvious. The rules of thumb are less obvious but equally important (see, e.g., Lenat, 1982). In physical domains they include empirical knowledge about the circumstances under which certain processes can be ignored (such as evaporation when water is poured from one glass to another) and what their net effect is (such as Black's law for the temperature of mixtures). Finally, different ontologies are sometimes necessary to deal with certain types of complex systems. In the process-oriented physics discussed here, describing flow requires finding flow paths. Finding flow paths in complex networks such as electrical circuits can quickly become computationally intractable; switching to a device-centered physics such as that described in deKleer and Brown (1983) can reduce the computational burden to manageable proportions for such systems.

In the balance beam example, it is known that the force of a block on the beam is qualitatively proportional to the mass of the block and to the distance from the fulcrum. If it is also known that the torque is the product of distance and weight, then providing numerical values for these quantities will allow an unambiguous prediction about tilt.

12.5.4.1 Learning Expert Models

The transition to expert models involves several kinds of learning. Some aspects of this transition probably lie outside the scope of experiential learning; for example, people typically learn mathematical models by being taught rather than by discovery. Some aspects of this learning—such as developing new ontologies—involve improving the content of the representations. Other aspects of the transition from a naive physics to an expert physics are better described as translating the existing qualitative representations into quantitative statements, using mathematics to express laws. By converting a physical theory into a mathematical model, the learner gains the ability to make precise predictions and to recognize powerful generalizations more easily. An important part of this refinement is to elaborate αQ statements into constraint equations. Langley (1979; Langley, Zytkow, Simon, and Bradshaw, 1983) describes techniques that should be useful for converting qualitative laws into mathematical relations.

Developing rules of thumb means knowing not just what is possible but what is probable. The learner must discover which outcomes raised by qualitative reasoning are likely or unlikely and which potential interactions can be ignored.

The techniques developed in machine learning for acquiring heuristics should be directly applicable (cf. Lenat, 1982; Mitchell et al., 1981). In addition, the authors suspect the possible behaviors raised by naive physics are compared against known protohistories. Hypothesized outcomes that have no corresponding protohistory are judged unlikely, and those corresponding to a highly familiar and accessible protohistory are judged very likely (see Tversky and Kahneman, 1973).

Further, it seems likely that at least some expert rules of thumb derive from learning new protohistories. This intuition is based in part on research in automaticity (Schneider and Fisk, 1983). It has been demonstrated that, given an orderly domain and sufficient practice, adult subjects can learn a new response pattern well enough so that it becomes essentially effortless (see also Anderson, 1982; Rumeihart and Norman, 1978). Moreover, there is some transfer from the learned material to new similar material. These learned sequences have many of the essential qualities of protohistories. First, they are triggered by recognition (in the terms used here, by a literal similarity match between the present situation and a stored situation). Second, computing and carrying out the procedures that follow from the match are automatic; virtually no attentional resources are required. Third, these computations are implicit; subjects are typically poor at introspecting about what they are doing, and when they do introspect, it can interfere with the response (Brooks, 1978; Reber, 1967, 1976). It may be too simplistic to view protohistories as a special case of automatic pattern-response combination, but there is enough overlap to allow some confidence that protohistories can continue to be learned at all stages of expertise. Of course, the *contents* of expert protohistories may be different from those of novices, since experts' protohistories may reflect a more advanced ontology, as discussed below. However, the mechanism of a perceptually triggered, automatic match should be the same.

The authors suspect that ontological shift is driven both by the desire to understand more complex physical systems and by the emergence of domain-independent mathematical abstractions. As an example of the first kind, consider the problem of reasoning about fluid flow in a complex system, such as a steam plant. Hayes (1979b) has distinguished two separate ontologies for liquids: a *contained-liquid* ontology, in which liquid is thought of as the fluid in a place, and a *molecular collection* ontology, in which water is thought of as little bits of fluid that move around inside the system. The contained-liquid ontology is appropriate if the goal is to determine what flows can occur. However, this view of water is not useful if one wants to know how changes the properties of the working fluid in one part of the system (say, the rising temperature of the inlet water in a boiler) affect properties of the fluid in another part of the system (say, temperature of the steam coming out of the boiler's superheater). In this case, liquid must be viewed in terms

of molecular collections that move around inside the system. Conversely, establishing flows using the molecular collection view is very difficult. A learner with only one of these two ontologies will have a difficult time with certain questions, and such difficulties may drive the search for a new Ontology.

Mathematical abstractions provide another important driving force in ontological change. In system dynamics, for example, physical systems involving fluid elements, mechanical elements, thermal elements, and acoustical elements are viewed as variations on a common, abstract theme. This means that the analysis and synthesis tools developed for abstract mathematical models can be used to solve problems in several domains. This is a powerful motivation, as evidenced by the wave of interest in attempting diverse applications evoked by the publication of certain new mathematical formalisms (e.g., catastrophe theory and fractal geometry).

12.6 SUMMARY

The authors have described their progress in weaving together Structure Mapping theory and Qualitative Process theory into a framework that aims to account for learning in physical domains. The learning sequence is built around three ideas. First, development proceeds from rich to sparse and from concrete to abstract—that is, initial representations differ from later representations in containing more, are more context-specific, information. Second, after sufficient experience people develop experiential models that are centered around the notion of physical process, as described by Qualitative Process theory. Third, the process of comparing and mapping between stored knowledge and the current situation, as described in Structure Mapping theory, is central to experiential learning.

Four stages of experiential learning have been laid out: protohistories, the causal corpus, naive physics, and expert models.¹⁶ The first stage, that of protohistories, embodies the idea that early representations are rich and context specific; this stage attempts to capture a combination of evidence from developmental patterns, similarity judgments, basic-level categories, and object prototypes. The third stage is the process-centered stage described by Qualitative Process theory. The fourth stage builds on the third-stage models, adding domain-independent generalizations and in some cases mathematical models. There is some

¹⁶In deriving this sequence of learning stages, the authors have been influenced by Piaget's well-known theory of cognitive development (see Piaget, 1954; or for an introduction to the work, see Flavel, 1963) However, the four stages of learning presented here differ considerably from Piaget's four-stage account. One difference, for example, is that the authors view their stages as domain specific, whereas Piaget stages are intended as general stages of intellectual development.

evidence for the third and fourth stages in the research on the novice-expert shift (Chi, Feltovich, and Glaser, 1981 Larkin, 1983).

The second stage, the causal corpus, is the most speculative. There is no direct evidence for its existence, nor do the authors currently have a detailed theory of the kinds of causal statements that can enter into the representations. Moreover, detailing how the causal corpus emerges from protohistories will not be easy. But something like the causal corpus seems necessary; a collection of simplistic, mostly binary directed regularities among dimensions and quantities that begin to be differentiated out of the tangled representations of the protohistory stage. The learner can now use these simple assertions as grist for further progress.

What happens to prior stages as new stages occur? First, stored representations have to be distinguished from new learning. The authors conjecture that learners retain much of their stored knowledge even when they go beyond the stage at which it was formed. Thus, a hydraulics engineer still uses the same protohistory he or she formed as a toddler to decide how fast one can carry a glass of water without spilling it. And, as deKleer points out (1979), expert physicists do not always resort to quantitative models (fourth stage); frequently the answer they want can be obtained by using a good qualitative model (third stage).

But what about new learning? Does new learning occur only at the leading edge, or do people continue to learn at levels below the most advanced stage they have attained? The authors suspect that even experts continue to learn at all prior stages, with the possible exception of the causal corpus. As described earlier, there is evidence that even experts continue to lay down new protohistories. Similarly, learners who are operating at the fourth stage, that of expert models, may continue to learn and refine their naive physics. This is because the mathematical models of the fourth stage are not a substitute for the process models of the third stage.¹⁷ Improvements to a naive physics are useful whether or not mathematical models are also available. As expertise increases the least new learning is expected within the causal corpus.

Of the four levels, the causal corpus has the least claim to continued independent existence in an advanced expert. The causal corpus is not reliable for prediction, nor does it possess the advantages of automaticity.¹⁸ In summary, the

¹⁷Historically, philosophers of science have differed about whether the best conception of a domain is provided by a mathematical model or by a mechanical model. For an extended discussion of this historical debate, see Hesse (1966). The position taken here is that both mathematical models and mechanical models are important to full understanding of a domain.

¹⁸This discussion, of course, concerns domains for which the learner eventually acquires expert knowledge. We suspect that people rely heavily on causal corpus knowledge in domains in which they are inexperienced. Further, there are many domains, such as child-rearing or getting rich, that lack definitive models. Collins's work on plausible reasoning (1978) suggests that in these domains, people rely heavily on this

overall picture is that a learner moves from rich, perceptual protohistories to the sparser representation of the causal corpus. The causal corpus serves as a staging area in which rough connections among variables can be stored until they can be subsumed into a true system. If learning continues, a person develops a process-centered naive physics and, for some domains, expert models.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the intellectual and financial support Schlumberger-Doll Research. They thank Allan Collins, Dave Chapman, Ed Smith Albert Stevens, and Dan Weld for their insightful comments on an earlier draft of the paper.

References

- Anderson, J. R., "Acquisition of Cognitive Skill," *Psychological Review*, Vol.89, No.4, pp. 369-404 1982.
- Brooks, L., "Nonanalytic Concept Formation and Memory for Instances," *Cognition and Categorization* B. B. Lloyd and E. Rosch (Eds.), Erlbaum, Hillsdale, N.J., 1978.
- Brown, A. L., and Campione, J. C., "Three Faces of Transfer: Implications for Early Competence, Individual Differences, and Instruction," in *Advances in Developmental Psychology*, M. Lam A. Brown, and B. Rogoff (Eds.), Erlbaum, Hillsdale, N.J., 1985, in press.
- Bruner, J. S., Olver, R., and Greenfield, P, *Studies in Cognitive Growth*, Wiley, New York, 1966.
Buckley, S., *Sun Up to Sun Down*, McGraw-Hill, New York, 1979.
- Burstein, M. H., "Concept Formation by Incremental Analogical Reasoning and Debugging," *Proceedings of the International Machine Learning Workshop*, R. S. Michalski (Ed.), Allerton House University of Illinois at Urbana-Champaign, pp.19-25, June 22-24, 1983. (An updated version this paper appears as chap. 13 of this volume.)
- Carbonell, J. O., "A Computational Model of Problem-Solving by Analogy," *Proceedings of the Seventh IJCAI*, Vancouver, B.C., pp.147-52, 1981.
- _____, "Derivational Analogy in Problem Solving and Knowledge Acquisition," *Proceedings of the International Machine Learning Workshop*, R. S. Michalski (Ed.), Allerton House,

causal corpus knowledge. See also Salter's work (1983) on tacit theories of economics.

- University Illinois at Urbana-Champaign, pp. 12-18, June 22-24, 1983. (An updated version of this paper appears as chap. 14 of this volume.)
- Carey, S., "Are Children Fundamentally Different Kinds of Thinkers and Learners Than Adults?" *Thinking and Learning Skills*, Vol.2, 5. Chipman, J. Segal, and R. Glaser (Eds.), Erlbaum, Hill dale, N.J., 1985, in press.
- Case, R., "Intellectual Development from Birth to Adulthood: A Neo-Piagetian Interpretation," in *Children's Thinking: What Develops?* R. S. Siegler (Ed.), Erlbaum, Hillsdale, N.J., 1978.
- Chi, M. T. H., Feltovich, P. J., and Glaser, R., "Categorization and Representation of Physics Problems by Experts and Novices," *Cognitive Science*, Vol.5, No.2, pp.121-51, April-June 1981.
- Clark, E. V., "What's in a Word? On the Child's Acquisition of Semantics in His First Language," in *Cognitive Development and the Acquisition of Language*, T. E. Moore (Ed.), Academic Press, New York, 1973.
- Clements, J., "Analogical Reasoning Patterns in Expert Problem-solving," *Proceedings of the Fourth Annual Conference of the Cognitive Science Society*, Berkeley, Calif., pp.137-40, 1982.
- Collins, A., "Fragments of a Theory of Human Plausible Reasoning," in *Theoretical Issues in Natural Language Processing II*, D. Waltz (Ed.), University of Illinois at Urbana-Champaign, 1978.
- Collins, A.; Warnock, E. H.; Aiello, N.; and Miller, M. L., "Reasoning from Incomplete Knowledge," in *Representation and Understanding: Studies in Cognitive Science*, D. Bobrow and A. Collins (Eds.), Academic Press, New York, 1975.
- Darden, L., "Reasoning by Analogy in Scientific Theory Construction," *Proceedings of the International Machine Learning Workshop*, R. S. Michalski (Ed.), Allerton House, University of Illinois at Urbana-Champaign, pp.3140, June 22-24, 1983.
- deKleer, J., "Causal and Teleological Reasoning in Circuit Recognition," Technical Report No.529, AI Lab, MIT, 1979.
- deKleer, J., and Brown, J. S., "Assumptions and Ambiguities in Mechanistic Mental Models," in *Mental Models*, D. Gentner and A. L. Stevens (Eds.), Erlbaum, Hillsdale, N.J., 1983.
- Dietterich, T., and Michalski, R. S., "A Comparative Review of Selected Methods for Learning from Examples," in *Machine Learning: An Artificial Intelligence Approach*, R. S. Michalski, J. O. Carbonell, and T. M. Mitchell (Eds.), Tioga, Palo Alto, Calif., 1983.
- diSessa, A., "Phenomenology and the Evolution of Intuition," in *Mental Models*, D. Gentner and

- A. L. Stevens (Eds.), Erlbaum, Hillsdale, N.J., 1983.
- Flavell, J. H., *The Developmental Psychology of Jean Piaget*, Van Nostrand Reinhold, New York, 1963.
- Forbus, K., "Qualitative Reasoning about Physical Processes," *Proceedings of the Seventh IJCAI*, Vancouver, B.C., pp.326-30, 1981.
- _____, "Measurement Interpretation in Qualitative Process Theory," *Proceedings of the Eighth IJCAI*, Karlsruhe, W. Ger., pp.315-20, 1983.
- _____, "Qualitative Process Theory," Technical Report No.789, AI Lab, MIT, 1984.
- Garner, W. R., "Aspects of Stimulus: Features, Dimensions, and Configurations," in *Cognition and Categorization*, B. B. Lloyd and E. Rosch (Eds.), Erlbaum, Hillsdale, N.J., 1978.
- Gelman, R., "Counting in the Preschooler: What Does and Does Not Develop," in *Children's Thinking: What Develops?* R. S. Siegler (Ed.), Erlbaum, Hillsdale, N.J., 1978.
- Gentner, D., "The Structure of Analogical Models in Science," Technical Report No.4451, Bolt Beranek and Newman, Cambridge, Mass., 1980.
- _____, "Are Scientific Analogies Metaphors?" in *Metaphor: Problems and Perspectives*, D. Miall (Ed.), Harvester Press, Ltd., Brighton, England, 1982.
- _____, "Structure-Mapping: A Theoretical Framework for Analogy," *Cognitive Science*, Vol.7, No.2, pp.155-70, 1983.
- Gentner, D., and Gentner, D. R., "Flowing Waters or Teeming Crowds: Mental Models of Electricity," in *Mental Models*, D. Gentner and A. L. Stevens (Eds.), Erlbaum, Hillsdale, N.J., 1983.
- Gentner, D., and Stuart, P., "Metaphor as Structure-Mapping: What Develops," Technical Report No.315, Center for the Study of Reading, University of Illinois, 1984.
- Gibson, E. J., *Principles of Perceptual Learning and Development*, Prentice-Hall, Englewood Cliffs, N.J., 1969.
- Gick, M. L., and Holyoak, K. J., "Analogical Problem Solving," *Cognitive Psychology*, Vol. 12 pp.306-55, 1980.
- _____, "Schema Induction and Analogical Transfer," *Cognitive Psychology*, Vol.15, pp.1-38, 1983.
- Hayes, P. J., "The Naive Physics Manifesto," in *Expert Systems in the Microelectronic Age*, D.

- Michie (Ed.), Edinburgh University Press, Edinburgh, 1979a.
- _____, "Naive Physics 1: Ontology for Liquids," Memo, Centre pour les etudes semantiques et cognitives, Geneva, Switzerland, 1979b.
- Hesse, M. B., *Models and Analogies in Science*, University of Notre Dame Press, Notre Dame, Md., 1966.
- Hoffman, R. R., "Metaphor in Science," in *The Psycholinguistics of Figurative Language*, R. P. Honeck and R. R. Hoffman (Eds.), Erlbaum, Hillsdale, N.J., 1980.
- Inhelder, B., and Piaget, J., *The Growth of Logical Thinking from Childhood to Adolescence*, Basic Books, New York, 1958.
- Kahneman, D., and Tversky, A., "Subjective Probability: A Judgment of Representitiveness," *Cognitive Psychology*, Vol.3, pp.430-54, 1972.
- Kempton, W., "Two Theories Used for Home Heat Control," in *Cultural Models in Language and Thought*, D. Holland and N. Quinn (Eds.), Cambridge University Press, Cambridge, 1985, in press.
- Langley, P., "Rediscovering Physics with BACON.3," *Proceedings of the Sixth IJCAI*, Tokyo, pp.505-7, 1979.
- Langley, P.; Zytkow, J.; Simon, H.; and Bradshaw, O., "Mechanisms for Qualitative and Quantitative Discovery," *Proceedings of the International Machine Learning Workshop*, R. S. Michalski (Ed.), Allerton House, University of Illinois at Urbana-Champaign, June 22-24, 1983. (An updated version of this paper appears as chap. 16 of this volume.)
- Larkin, J., "The Role of Problem Representation in Physics," in *Mental Models*, D. Gentner and A. L. Stevens (Eds.), Erlbaum, Hillsdale, N.J., 1983.
- Lenat, D., "The Nature of Heuristics," *Artificial Intelligence*, Vol.19, pp.189-249, 1982.
- _____, "The Nature of Heuristics II," *Artificial Intelligence*, Vol.20, pp.31-5⁹, 1983. "The Nature of Heuristics III," *Artificial Intelligence*, Vol.20, pp.61-98, 1983.
- McCarthy, J., and Hayes, P., "Some Philosophical Problems from the Standpoint of Artificial Intelligence," in *Machine Intelligence 4*, Edinburgh University Press, Edinburgh; 1969.
- Michalski, R. S., "A Theory and Methodology of Inductive Learning," *Artificial Intelligence*, Vol 2 pp.111-61, 1983.
- Michalski, R. S., and Stepp, R. E., "Learning from Observation: Conceptual Clustering," in *Machine Learning: An Artificial Intelligence Approach*, R. S. Michalski, J. O. Carbonell, and T. M. Mitchell (Eds.), Tioga, Palo Alto, Calif., 1983.

- Mitchell, T. M., "Generalization as Search," *Artificial Intelligence*, Vol.18, pp.203-26, 1982.
- Mitchell, T. M.; Utgoff, P. E.; Nudel, B.; and Banerji, R., "Learning Problem-solving Heuristics Through Practice," *Proceedings of the Seventh IJCAI*, Vancouver, B.C., pp. 127-34, 1981.
- Nelson, K., "How Children Represent Knowledge of Their World in and out of Language: A Preliminary Report," in *Children's Thinking: What Develops?* R. S. Siegler (Ed.), Erlbaum, Hillsdale, N.J., 1978.
- Nickerson, R. S., and Adams, M. J., "Long-term Memory for a Common Object," *Cognitive Psychology*, Vol.11, pp. 287-307, 1979.
- Palmer, S. E., "Fundamental Aspects of Cognitive Representation," in *Cognition and Categorization*, B. B. Lloyd and E. Rosch (Eds.), Erlbaum, Hillsdale, N.J., 1978.
- Piaget, J., *The Child's Conception of Physical Causality*, Routledge & Kegan Paul, London, 1951. *The Construction of Reality in the Child*, Basic Books, New York, 1954.
- Posner, M. I., and Mitchell, R. F., "Chronometric Analysis of Classification," *Psychological Review*, Vol.74, pp.392-409, 1967.
- Reber, A. S., "Implicit Learning of Artificial Grammars," *Journal of Verbal Learning and Verbal Behavior*, Vol. 6, pp. 855-63, 1967.
- _____, "Implicit Learning of Synthetic Languages: The Role of Instructional Set," *Journal of Experimental Psychology: Human Memory and Learning*, Vol.2, pp. 88-94, 1976.
- Reed, S. K., Ernst, O. W., and Banerji, R., "The Role of Analogy in Transfer Between Similar Problem States," *Cognitive Psychology*, Vol.6, pp.436-50, 1974.
- Rosch, E., "On the Internal Structure of Perceptual and Semantic Categories," in *Cognitive Development and the Acquisition of Language*, T. E. Moore (Ed.), Academic Press, New York, 1973.
- _____, "Cognitive Representations of Semantic Categories," *Journal of Experimental Psychology: General*, Vol. 107, pp. 192-233, 1975.
- _____, "Principles of Categorization," in *Cognition and Categorization*, E. Rosch and B. B. Lloyd (Eds.), Erlbaum, Hillsdale, N.J., 1978.
- Rumelhart, D., and Norman, D. A., "Accretion, Thning, and Restructuring: Three Modes of Learning," in *Semantic Factors in Cognition*, J. W. Cotton and R. L. Klatzky (Eds.), Erlbaum, Hillsdale, N.J.,

- Schneider, W., and Fisk, A. D., "Attention Theory and Mechanisms for Skilled Performance," in *Memory and Control of Action*, R. A. Magill (Ed.), North-Holland, New York, 1983.
- Shepp, B. E., "From Perceived Similarity to Dimensional Structure: A New Hypothesis about Perspective Development," in *Cognition and Categorization*, E. Rosch and B. B. Lloyd (Eds.), Erlbaum, Hillsdale, N.J., 1978.
- Siegler, R. S., "Three Aspects of Cognitive Development," *Cognitive Psychology*, Vol.8, pp. 481-520,
- _____, "The Origins of Scientific Reasoning," in *Children's Thinking: What Develops?* in R. S. Siegler (Ed.), Erlbaum, Hillsdale, N.J., 1978.
- _____, "Developmental Sequences Within and Between Concepts," *Monographs of the Society for Research in Child Development*, Vol.46, No.2, 1981.
- Smith, E. E., and Medin, D. L., *Categories and Concepts*, Harvard University Press, Cambridge, 1981.
- Stepp, R. E., and Michalski, R. S., "How to Structure Structured Objects," *Proceedings of the International Machine Learning Workshop*, R. S. Michalski (Ed.), Allerton House, University of Illinois at Urbana-Champaign, pp.156-60, June 22-24, 1983. (An updated version of this paper appears as chap. 17 of this volume.)
- Stevens, A., Collins, A., and Goldin, S. E., "Misconceptions in Students' Understanding," *Journal of Man-Machine Studies*, Vol.11, pp.145-56, 1979.
- Sussman, G. J., *A Computational Model of Skill Acquisition*, Springer-Verlag, New York, 1976.
- Tversky, A., "Features of Similarity," *Psychological Review*, Vol.84, pp.327-52, 1977.
- Tversky, A., and Gati, I., "Similarity, Separability and the Triangle Inequality," *Psychological Review*, Vol.89, No.2, pp.123-54, 1982.
- Tversky, A., and Kahneman, D., "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology*, Vol.5, pp.207-32, 1973.
- Williams, M., Hollan, J., and Stevens, A. L., "Human Reasoning about a Simple Physical System," in *Mental Models*, D. Gentner and A. L. Stevens (Eds.), Erlbaum, Hillsdale, N.J., 1983.
- Winner, E., "New Names for Old Things," *Journal of Child Language*, Vol.6, pp.469-91, 1979.
- Winston, P. H., "Learning and Reasoning by Analogy," *Communications of the ACM*, Vol.23, No.12, pp. 689-703, 1980.

_____, "Learning New Principles from Precedents and Exercises," *Artificial Intelligence*, Vol. 19, pp.321-50, 1982.