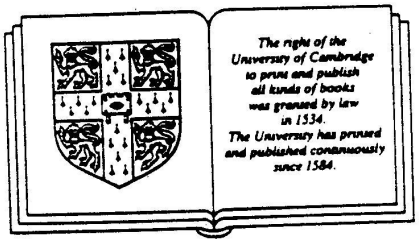
89-30

Similarity and analogical reasoning

Edited by

STELLA VOSNIADOU
ANDREW ORTONY



CAMBRIDGE UNIVERSITY PRESS
Cambridge
New York *New Rochelle* *Melbourne* *Sydney*

4

Two-tiered concept meaning, inferential matching, and conceptual cohesiveness

RYSZARD S. MICHALSKI

Introduction

Suppose we asked someone how to get to some place in the city we were visiting and received needed instructions in response. Clearly, we would say that this person *knew* the answer, no matter whether the person knew the place personally or just had to figure out its location on the basis of general knowledge of the city, that is, by conducting inference. We would say this, of course, only if the answer were given to us in a reasonable amount of time.

The above example illustrates a general principle: One knows what one remembers, or what one can infer from what one remembers within a certain time constraint. Thus our knowledge can be viewed as a combination of two components, memorized knowledge and inferential extension, that is, knowledge that can be created from recorded knowledge by conducting inference within a certain time limit.

The main thesis of this chapter is that individual concepts – elementary components of our knowledge – parallel such a two-tiered nature of knowledge. We hypothesize that processes of assigning meaning to individual concepts recognized in a stream of information, or of retrieving them from memory to express an intended meaning are intrinsically inferential and involve, on a smaller scale, the same types of inference – deductive, analogical, and inductive – as processes of applying and constructing knowledge in general. This hypothesis reflects an intuition that the meaning of most concepts cannot, in principle, be defined in a crisp and context-independent fashion.

Specifically, the meaning of most concepts cannot be completely defined by some necessary or sufficient features, by a prototype, or by a set of representative exemplars. Rather, the meaning of a concept is a dynamic structure built each time anew, in the course of an interaction between some initial base meaning and the interpreter's background knowledge in the given context of discourse.

This view leads us to the proposition that the meaning we assign to a concept in any given situation is a result of an interplay between two parts: the *base concept representation* (BCR), and the *inferential concept interpretation* (ICI). The base concept representation is an explicit structure residing in memory that records both specific facts about the concept and general characteristics of it. The specific facts may include representative examples, exceptions. and counterexamples. The general characteristics are teacher-defined, or inferred by induction from examples or by analogy. They include typical, easily definable, and possibly context-independent assertions about the concept. These characteristics tend to capture the principle, the ideal or intention behind a given concept. If this principle changes to reflect a deeper knowledge about the concept involved, the BCR is redefined. To see this, consider, for example, the changes of our understanding of concepts such as *whale* (from fish to mammal) or *atom* (from the smallest indivisible particle to the contemporary notion of a dual wave–matter form).

The inferential concept interpretation is a process of assigning meaning to a concept using the BCR and the context of discourse. This process involves the interpreter's relevant background knowledge and inference methods and transformations that allow one to recognize, extend, or modify the concept meaning according to the context. These methods are associated with the concept or its generalizations, and, together with relevant background knowledge, constitute the second tier in concept representation.

The main goal of this chapter is to sketch ideas and underlying principles for constructing an adequate cognitive model of human concepts. It is not to define such a model precisely or to present specific algorithms. It is also hoped that the proposed ideas will suggest better computational methods for representing, using, and learning concepts in artificial intelligence systems.

Inference allows us to remember less and know more

This section will attempt to show that the two-tiered representation of concept meaning outlined above can be justified on the basis of cognitive economy – that is, economy of mental resources, memory and processing power – and that it reflects some general aspects of the organization of human memory. For a discussion of issues concerning cognitive economy see Lenat, Hayes-Roth, and Klahr (1979).

Let us start by assuming that the primary function of our knowledge is to interpret the present and predict the future. When one is exposed

to any sensory inputs, one needs knowledge to interpret them. The more knowledge and the stronger the inferential capabilities (i.e., roughly the number of production and inference rules) one possesses, the greater the amount of information one can derive from a given input.

Interpreting observations in the context of the available knowledge makes it possible to derive more information from the input than is presented on the surface. It also allows one to build expectations about the results of any action and to predict and/or influence future events. The latter is possible because events and objects in our world are highly interrelated. If our world consisted of totally unrelated random events, one following the other, our knowledge of the past would be of no use for predicting the future, and this would obviate the need to store any knowledge. Moreoever, this, in turn, would presumably obviate the need for having intelligence, as the primary function of intelligence is to construct and use knowledge.

On the other hand, if our world were an eternal repetition of exactly the same scenes and events, knowledge once acquired would be applicable forever, and the need for its extension and generalization would cease. No wonder that in old, slow-changing traditional societies the elderly enjoyed such high status. The slower the rate of change in an environment, the higher the predictive value of past specific knowledge and the lower the need to extend and generalize knowledge. This suggests a hypothesis that the degree to which our innate subconscious capabilities for generalizing any input information corresponds to the rates of change in our environment. Thus, in a world that was evolving and changing at a different rate, our innate capabilities for generalization would presumably be different.

From the myriad sensory inputs and deluge of information received, we select and store only a minuscule fraction. This selection is done by a goal-dependent filtering of the inputs. The fraction actually stored contains a spectrum of structures representing different levels of abstraction from reality and different beliefs in their correctness. This spectrum spans the low-level, highly believed facts and observations, through partial plausible abstractions and heuristics, to high-level and highly hypothetical abstractions. The highest belief usually is assigned to our own personal sensory experiences, and the lowest belief to vague abstractions made by people whom we do not especially trust. These various assertions, together with a degree of belief in them, are automatically memorized when they are received or generated by inference. They then can be forgotten but not consciously erased.

The filtering of input information is done by conducting inferences – deductive, analogical, and inductive – that engage the input information and the goals and the knowledge of the person. The idea that a person's knowledge is involved in the processes of interpreting inputs is, of course, not new. An interesting illustration of it is presented, for example, by Anderson and Ortony (1975). They conducted experiments showing that the comprehension of a sentence depends heavily on the person's knowledge of the world and his or her analysis of the context.

Our ability to make inferences seems to come from a naturally endowed mechanism that is automatically activated in response to any input of information. One may ask why this is so. As our memory and information-processing powers are limited, it seems natural that the mind should tend to minimize the amount of information stored and maximize the use of that which is already stored. Consequently, one may hypothesize that the inferential processes that transfer any input information to stored knowledge are affected by three factors:

1. what is important to one's goals
2. what knowledge will be maximally predictive
3. what knowledge will allow one to infer the maximum amount of other knowledge.

The first factor reflects the known phenomenon that facts considered very important tend to be remembered before other facts. The second factor is significant because the predictive power of knowledge enables us to develop expectations about the future, and thus to prevent or avoid undesirable courses of actions, and to achieve goals. The third factor relates to cognitive economy: If we can infer *B* from *A* without much cognitive effort, then it is enough just to remember *A*. The second and third factors have interesting consequences. They suggest a memory organization that is primarily oriented toward storing analogies and generalizations, but facilitates the process of efficiently performing deduction on the knowledge stored.

These three factors explain the critical role of analogical and inductive inference in the process of transforming information received from the environment to knowledge actually memorized. This is so because it is analogical inference that transfers knowledge from known objects or problem solutions to new but related objects or problems. And it is inductive inference that produces generalizations and causal explanations of given facts (from which one can deduce original facts and predict new ones). Strict deductive inference and various forms of plausible inference (plausible deductive, analogical,

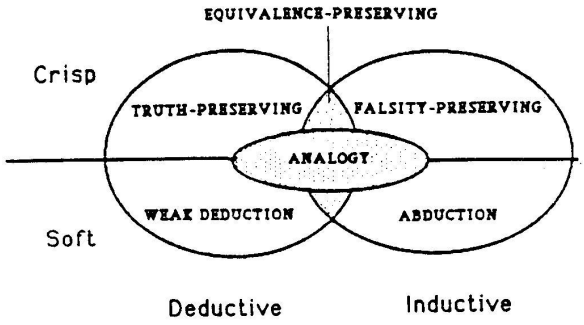


Figure 4.1. Types of inference.

and inductive) are means for extending/deriving more knowledge from our base knowledge, though such derived knowledge may be of lesser certainty.

The relationship between different types of inference is shown in Figure 4.1. The types of inference are divided according to two dimensions: (a) mode of inference: deductive versus inductive; and (b) strength of inference: crisp versus plausible. "Crisp" deductive inference is the truth-preserving inference studied in formal logic. "Soft" deductive inference uses approximate rules of deductive inference and produces probable rather than strict consequences of given premises. This type of inference is implemented, for example, in various expert systems that generate advice together with an estimate of its certainty.

Inductive inference produces hypotheses (or explanations) that crisply or softly imply original facts (premises). This means that original facts are deductive consequences of the hypotheses. Crisp inductive inference is a falsity-preserving inference. For example, hypothesizing that all professors at a particular university are bright on the basis that all professors of the Computer Science Department at this university are bright is a falsity-preserving inductive inference. (If the initial premise is true, the conclusion can be true or false; but if the premise is false, the hypothesis must be false also. Conversely, if the hypothesis is true, then the premise clearly must be true also.) Soft inductive inference produces hypotheses that only plausibly imply the original facts. For example, seeing smoke, one may hypothesize that there is a fire somewhere. It is a soft inductive inference, because there could be smoke without a fire.

Analogical inference is placed in the middle because it can be viewed

as inductive and deductive inference combined (Michalski, 1987). The process of noticing analogy and creating an analogical mapping between two systems is intrinsically inductive; the process of deriving inferences about the analog using the mapping is deductive. This view, derived by the author through purely theoretical speculations, seems to be confirmed by the experimental findings of Gentner and Landers (1985) and Gentner (this volume). In order to explain difficulties people have in noticing analogies, they decomposed analogical reasoning into three parts, which they call "access," "structure-mapping," and "inferential power." They found that access and structure-mapping are governed by different rules than inferential power. Access is facilitated by literal similarity or mere appearance, and structure-mapping is governed by similarity of higher-order relations. These are inductive processes, as they produce a structure that unifies the base and the target systems. Inferential power corresponds to deduction.

The view of analogy as induction and deduction combined explains why it is more difficult for people to notice analogy than to use it once it is observed. This is so because inductive inference, being an underconstrained problem, typically consumes significantly more cognitive power than deductive inference, which is a well-constrained problem.

Figure 4.2 illustrates levels of knowledge derived from the base knowledge by conducting various types of inference (the "trumpet" model). The higher the type of inference, the more conclusions can be generated, but the certainty of conclusions decreases. A core theory and a discussion of various aspects of human plausible inference can be found in Collins and Michalski (1986).

Let us now return to the discussion of the third factor influencing inferential processes, namely, what knowledge allows us to infer the maximum amount of other knowledge. This issue, obviously, has special significance for achieving cognitive economy. The need for cognitive economy implies that it is useful for individual words (concepts) to carry more than one meaning, when considered without any context and without inferential extension of their meaning. By allowing that the meaning of words can be context-dependent and inferentially extensible, one can greatly expand the number of meanings that can be conveyed by individual words. This context dependence, however, cannot be unlimited, again because of cognitive economy. To be economical, context dependence should be employed only when the context can be identified with little mental effort. Inferential extensions

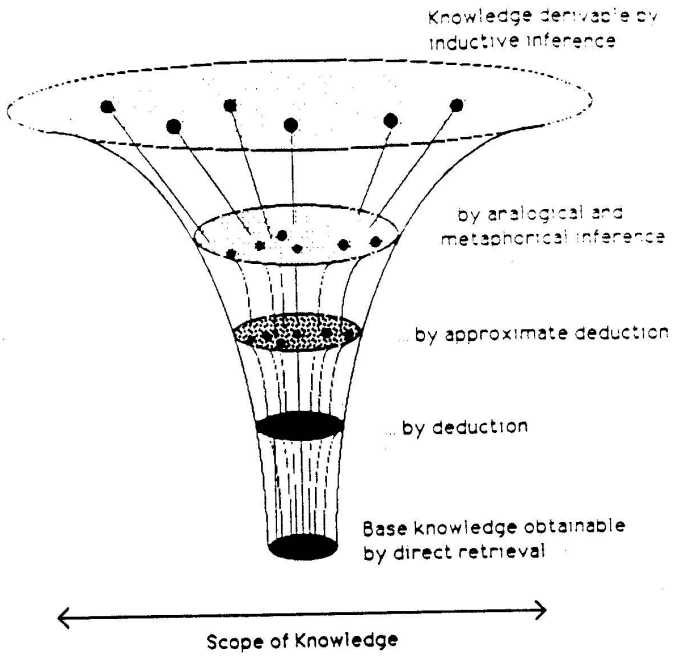


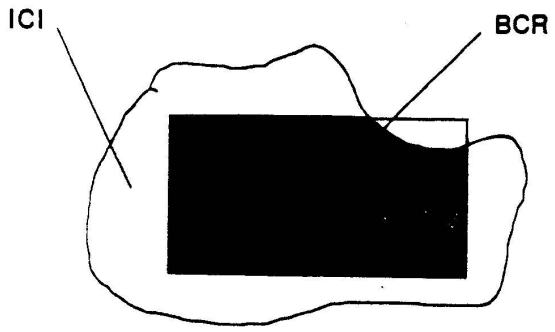
Figure 4.2. A "trumpet" model of inferential knowledge extension. Shading represents decreasing strength of belief in inferentially derived knowledge.

also have natural limits, which are dictated by the mental power available, and the decreasing confidence in conclusions as the levels of inference increase.

Concept meaning is distributed between representation and interpretation

Concepts are mental structures representing classes of entities united by some principle. Such a principle might be a common use or goal, the same origin or behavior, or just similar perceptual characteristics. In order to use concepts, one must possess efficient methods for recognizing them in streams of sensory signals or in mental processing. To do so, one needs to have appropriate mental representations of concepts.

The traditional work on concept representation assumes that the whole meaning of a concept resides in a single stored structure, for example, a semantic network, a frame, or a graph, that captures all relevant properties of the concept (e.g., Collins & Quillian, 1972;



BCR - the scope of the concept defined by the Base Concept Representation

ICI - the scope of the concept as derived by Inferential Concept Interpretation for a given context and background knowledge

Figure 4.3. An illustration of two-tiered concept representation.

record, fingerprints, any combination of these, or by a host of other characteristics. Thus, if the concept recognition process were based on a direct match of a fixed number of features of the target concept with properties of an observed object, then one would need to store representations for all these possibilities. Such a method would be hopelessly memory-taxing and inefficient. It is practical only in simple cases, such as those considered in many current expert systems.

In the proposed theory, the process of relating the base representation of a concept to observations is done by inferential concept interpretation. This process "matches" the base concept representation with observations by conducting inference involving the contextual information (e.g., What are other candidate concepts?) and relevant background knowledge. This inference determines what features are needed or sufficient to be matched in order to recognize a concept among a context-dependent set of candidates, and what kind of match is required. Thus the degree of match between a concept representation (CR) and an observed entity (OE) is not just a function of CR and OE, as traditionally assumed, but rather a four-argument function, which also includes a parameter, CX, for context, and BK, for background knowledge:

$$\text{Degree of match (CR, OE)} = f(\text{CR, OE, CX, BK})$$

The context is computed dynamically in the process of using or recognizing concepts. Thus the proposed view requires an efficient method for representing and using contexts for any given concept. A simple introspection of our mental processes appears to confirm this: We seem to have little difficulty in determining and maintaining the context in any discourse.

There is no unique way of distributing the concept meaning between BCR and ICI. We expect that the actual distribution of the concept meaning between these two parts represents a desired tradeoff between the economy of concept representation and the economy of inferential concept interpretation. Thus, learning a concept involves acquiring not only the base concept representation but also the methods for inferential concept interpretation.

Let us illustrate the proposed approach by a few examples. Consider the concept of *fish*. Typical and general characteristics of fish are that they have a certain elongated shape, a tail, live in water, and swim. These and other typical physical properties of fish, as well as representative examples, would be stored in the BCR. Suppose someone finds an animal that matches many characteristics of fish but does not swim. Suppose that this animal appears to be sick. The ICI would involve background knowledge that sick animals may not be able to move and that swimming is a form of moving. By deductive reasoning from these facts one concludes that lack of ability to swim should not be taken as negative evidence for the animal being a fish. On the contrary, the fact that the animal does not swim might even add to the confidence that it is a fish, once the animal was recognized as being sick.

Suppose that we learned the concept of *fish* by reading a general description and seeing a few examples. The BCR consists of this general description and the memorized examples. Suppose that we visit a zoo and see an animal defined as *fish* that is of a shape never seen in the examples or stated in the general description – say, a horselike shape. We may add this example to our BCR without necessarily modifying our general notion of *fish*. If we see another horse-shaped fish, we may recall that example and recognize the new instance as a fish without evoking the general notion of *fish*. This explains why we postulate that the BCR is not just a representation of the general, typical, or essential meaning of a concept but may also include examples of a concept.

The rules used in the above reasoning about sick fish would not be stored as the base concept representation for *fish*. They would be a part of the methods for inferential concept interpretation. These

methods would be associated with the general concept of *animal*, rather than with the concept of *fish*, because they apply to all animals. Thus we postulate that the methods for inferentially interpreting a concept can be inherited from those applicable to a more general concept.

As another example, consider the concept of *sugar maple*. Our prototypical image of a sugar maple is that it is a tree with three- to-five-lobed leaves that have V-shaped clefts. Some of us may also remember that the teeth on the leaves are coarser than those of the red maple, that slender twigs turn brown, and that the buds are brown and sharp-pointed. Being a tree, a sugar maple has, of course, a trunk, roots, and branches.

Suppose now that while strolling on a nice winter day someone tells us that a particular tree is a sugar maple. Simple introspection tells us that the fact that the tree does not have leaves would not strike us as a contradiction of our knowledge about sugar maples. This is surprising, because, clearly, the presence of leaves of a particular type is deeply embedded in our typical image of a maple tree. The two-tiered theory of concept representation explains this phenomenon simply: The inferential concept interpretation associated with the general concept of *tree* evokes a rule; "In winter deciduous trees lose leaves." By deduction based on the subset relationship between a tree and a maple tree, the rule would be applied to the latter. The result of this inference would override the stored standard information about maple trees, and the inconsistency would be resolved.

Suppose further that when reading a book on artificial intelligence we encounter a drawing of an acyclic graph structure of points and straight lines connecting them, which the author calls a tree. Again, calling such a structure a tree does not evoke in us any strong objection, because we can see in it some abstracted features of a tree. Here, the matching process involves inductive generalization of the base concept representation. Once such a generalized notion of a tree is learned in the context of mathematical concepts, it will be used in this context.

These examples clearly show that the process of relating observations with concept representations is much more than matching features and determining a numerical score characterizing the match, as done in various mechanized decision processes, for example, expert systems.

It should be noted that the distribution of the concept meaning between the representation and interpretation parts is not fixed but can be done in many ways. Each way represents a trade-off between

the amount of memory for concept storage and computational complexity of concept use. At one extreme, all the meaning can be expressed by the representation. In this case the representation explicitly defines all properties of a concept, including any concept variations, exceptions, and irregularities. It states directly the meaning of the concept in every possible context. It stores all known examples of the concept. This results in a very complex and memory-taxing concept representation. The concept interpretation process would, however, be relatively simple. It would involve a straightforward matching of the properties of the unknown object with information in the concept description.

At the other extreme, the concept is explicitly represented only by the most simple description characterizing its idealized form. In this case, the process of matching a concept description with observations might be significantly more complex.

As far as memory representation of concepts is concerned, we assume that their base concept representations are stored as a collection of assertions and facts. These collections are organized into *part* or *type* hierarchies with inheritance properties (e.g., Collins & Michalski, 1986). The methods used by inferential concept interpretation are also arranged into hierarchies. For example, as already indicated, the rule that a sick fish may not swim is stored not with the ICI methods associated with the concept of *fish* but rather with the concept of *animal*.

As mentioned earlier, the process of inferential concept interpretation may involve performing not just truth-preserving deductive inference on the base concept interpretation but also various forms of plausible inference. In particular, it may create an inductive generalization of the base concept representation, draw analogies, run mental simulations, or envision consequences of some acts or features. The background knowledge needed for inferential interpretation includes information about methods for relating concept representations to observations, about which properties are important and which are not in various contexts, and about typicality of features, statistical distribution of properties and concept occurrences, and so on. An inferential interpreter may produce a yes/no answer or a score representing the degree to which the base representation matches given observations. Extending the meaning of a single concept by conducting inference corresponds on a small scale to extending any knowledge by inference.

When an unknown entity is matched against a base concept representation, it may satisfy it directly or it may satisfy some of its in-

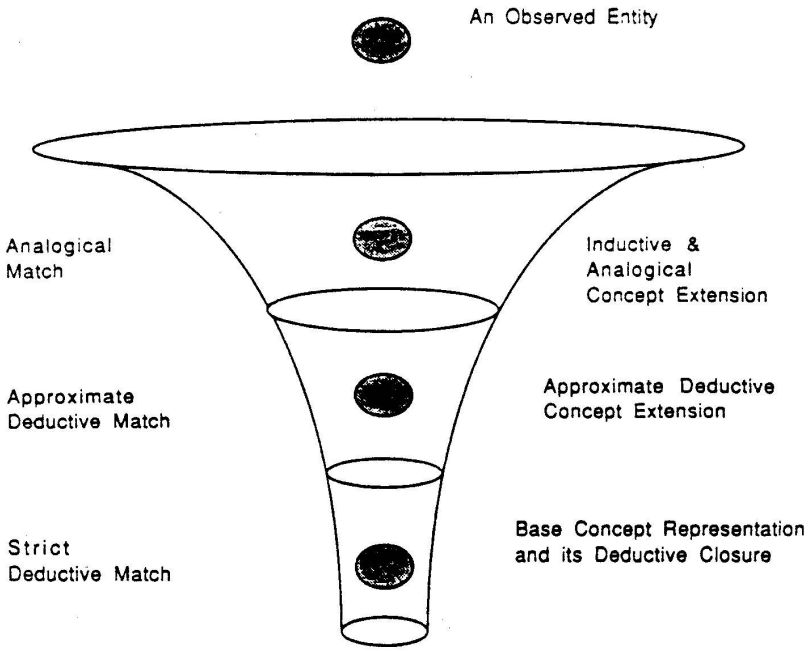


Figure 4.4. Types of inferential concept matching.

ferential extensions. The type of inference performed to match the description of the entity with the base concept representation determines the type of match (Figure 4.4). If the description of an entity matches the BCR precisely – satisfies it directly or satisfies its specialization (falls into its deductive extension) – then we have a *strict match*; if it satisfies an approximate deductive extension, then we have an *approximate match*; if it matches an analogical or inductive extension – satisfies a generalization that unifies the BCR with the description of the entity – then we have an *analogical* or, generally, an inferential match.[1]

As mentioned earlier, when we are recognizing an entity in the context of a finite set of candidate entities, usually only a small subset of its properties will need to match the properties in the base representation of candidate concepts. This set is defined by a *discriminant concept description* (Michalski, 1983). Such a description can be determined by conducting inductive inference on the base representation of the candidate concepts. A method for an efficient recognition of concepts in the context of candidate concepts, called *dynamic recognition*, is described in Michalski (1988).

The process of inferential concept interpretation can be viewed as a vehicle for extending the base concept meaning into a large space of variations by the use of context, rules of inference, and general knowledge. This process is an important means for achieving flexibility of concepts and thus leads to cognitive economy. Later, in the section describing experimental results, we present an example of a very simple inferential interpretation of a logic-style base concept representation.

Some other views on concept representation

There seems to be universal agreement that human concepts, except for special cases occurring predominantly in science (concepts such as a prime number, a triangle, a vertebrate, etc.), are structures with flexible and imprecise boundaries. I call such concepts *flexible*. They allow a varying degree of match between them and observed instances and have context-dependent meaning. Flexible boundaries make it possible to "fit" the meaning of a concept to changing situations and to avoid precision when it is not needed or not possible. The varying degree of match reflects the varying representativeness of a concept by different instances. According to the theory presented, this is accomplished by applying inferential concept matching, which takes into consideration the context and background knowledge of the interpreter.

Instances of a concept are rarely homogeneous. Among instances of a concept people usually distinguish a "typical instance," a "nontypical instance," or, generally, they rank instances according to their typicality. By using context, the meaning of almost any concept can be expanded in directions that cannot be predicted in advance. An illustration of this is given by Hofstadter (1985, chap. 24), who shows how a seemingly well-defined concept, such as *First Lady*, can express a great variety of meanings depending on the context. For example, it might include the husband of Margaret Thatcher.

Despite various efforts, the issue of how to represent concepts in such a rich and context-dependent sense is not resolved. Smith and Medin (1981) distinguish among three approaches: the *classical view*, the *probabilistic view*, and the *exemplar view*. The classical view assumes that concepts are representable by features that are singly necessary and jointly sufficient to define a concept. This view seems to apply only to very simple cases. The probabilistic view represents concepts as weighted, additive combinations of features. It postulates that concepts should correspond to linearly separable subareas in a feature

space. Experiments indicate, however, that this view is also not adequate (Smith & Medin, 1981; Wattenmaker, Dewey, Murphy, & Medin, 1986). The exemplar view represents concepts by one or more typical exemplars rather than by generalized descriptions. Although it is easy to demonstrate that we do store and use concept exemplars for some particular purposes, it seems clear that we also create and use abstract concept representations. For important ideas on concept representation and organization from the computational viewpoint, see papers by Minsky (1980), Sowa (1984), and Lenat, Prakash, and Shepherd (1986).

The notion of typicality can be captured by a measure called *family resemblance* (Rosch & Mervis, 1975). This measure represents a combination of frequencies in which different features occur in different subsets of a superordinate concept, such as *furniture, vehicle*, and so on. The individual subsets are represented by typical members. Nontypical members are viewed as corruptions of the typical, differing from them in various small aspects, as children differ from their parents (e.g., Rosch & Mervis, 1975; Wittgenstein, 1953). The idea of family resemblance is somewhat related to the two-tiered representation, except that the BCR is a much more general concept than a prototype, and the ICI represents a significantly greater set of transformations than "corruptions" of a prototype.

Another approach uses the notion of a *fuzzy set* as a formal model of imprecise concepts (Zadeh, 1976). Members of such a set are characterized by a graded set-membership function rather than by the in/out function employed in the classical notion of a set. This set-membership function is defined by people describing the concept and thus is subjective. This approach allows one to express explicitly the varying degree of membership of entities in a concept, as perceived by people, which can be useful for various applications. It does not explain, however, what are the computational processes that determine the set-membership functions. Neither is it concerned with developing adequate computational mechanisms for expressing, handling, and reasoning about the context-dependence and background-knowledge dependence of the concept meaning.

The idea of two-tiered representation attributes the graded concept membership to the flexibility of inferential concept interpretation. Thus, instead of explicitly storing the membership function, one obtains an equivalent result as a by-product of the method of interpreting the base concept representation. This method also handles the context and background-knowledge dependence, via a store of rules and allowable concept transformations.

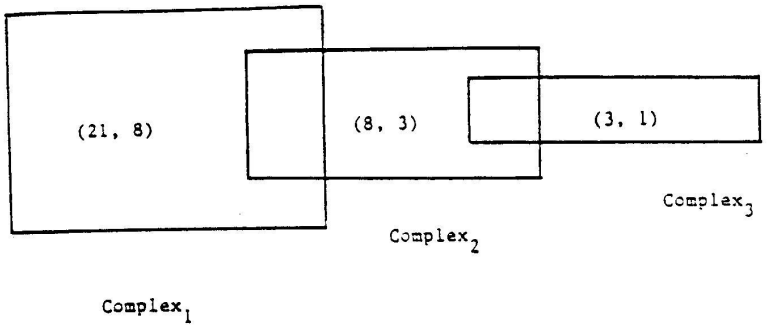


Figure 4.6. An ordered disjunctive concept representation. The numbers in parentheses denote the t weight and u weight, respectively.

the blood type A or O occurred in a total of 60 patients with this disease, and in 55 patients it occurred uniquely (i.e., these patients did not have properties satisfying other complexes associated with this disease). Statements with high t weights may be viewed as characterizing typical cases of a concept, and statements with low t weights and u weights can be viewed as characterizing rare, exceptional cases, or errors.

In the experiment, initial covers for the diseases were determined by applying the inductive learning program AQ15 to a set of known cases of diseases (Hong, Mozetic, & Michalski, 1986; Michalski, Mozetic, Hong, & Lavrac, 1986). Complexes in each cover were ordered according to decreasing values of t weights, as shown in Figure 4.6. (If two complexes had the same t weight, then they were ordered by decreasing values of u weights). Thus, the first complex in each cover is likely to characterize the most typical properties of the concept, the next complex less typical properties, and so on.

As mentioned earlier, a cover serves here the role of the base concept representation. The diagnosis of a case is determined by matching each cover with the case, and finding the cover with the highest match. The way the cover is matched against a disease case is determined by the inferential concept interpretation (see the discussion of flexible matching below).

To determine the most desirable distribution of the concept meaning between the BCR and ICI, the so-called TRUNC method was applied.

First, the initial covers for each disease obtained from AQ15 were used to diagnose a set of new disease cases, and the performance score was calculated. Next, each such cover was reduced by removing from it the "lightest" complex (i.e., the complex with the smallest t weight).

are presented in the work by DeJong (1986), DeJong and Mooney (1986), and Mitchell, Keller, and Kedar-Cabelli (1986). A computational framework for applying *plausible* inference for interpreting observations (specialization, generalization, and similarity-based transformations) is described by Collins and Michalski (1986). Various issues involved in creating mental representations of concepts are described by Collins and Gentner (1987).

The next section describes an experimental study investigating a simple form of two-tiered concept representation in the context of learning decision rules from examples in the area of medicine.

The two-tiered representation can reduce memory needed: an experiment

This section describes the results of an experiment investigating a simple form of two-tiered representation of four different types of lymphography. In the experiment, the base concept representation, called a *cover*, is in the form of a disjunction of conjunctive statements, called *complexes*. Interpreting a complex as the condition part of a rule, CONDITION → CONCEPT NAME, a cover can be viewed simply as a set of rules with the same right-hand side.

The complexes are conjunctions of relational assertions, called *selectors*. Each selector characterizes one aspect of the concept. It states a value or a set of values that an attribute may take on for the entities representing the concept. Here are two examples of selectors:

> [blood type = A or B] (*Read*: The blood type is A or B.)
> [Diastolic blood pressure = 65 . . . 90] (*Read*: The diastolic blood pressure is between 65 and 90.)

Thus selectors relate an attribute to one or more of the attribute's possible values. A selector is *satisfied* by an entity if the selector's attribute applied to this entity takes on one of the values stated in the selector. Each complex (a conjunction of selectors) in the base concept representation (cover) is associated with a pair of weights, t and u, representing, respectively, the *total* number of known cases that it covers and the number of cases that it covers alone *(uniquely)* among other complexes associated with this concept. For example, suppose that the complex:

> [blood pressure = 140/90] & [blood type = A or O]: 60, 55

is one of the complexes characterizing patients with some disease. The weights $t = 60$ and $u = 55$ mean that the blood pressure 140/90 and

Table 4.1. *Experimental results*

Domain	Cover reduction	Complexity # Sel	Complexity # Cpx	Accuracy 1st choice (percentage)	Human experts	Random choice
Lymphography	none	37	12	81	60–85% (estimate)	25%
	unique >1	34	10	80		
	best cpx	10	4	82		

individual complexes in a cover is summed as probabilities. Specifically, the degree of match is computed according to these rules:

- The degree of match (DM) between an event and a selector is 1 when the selector is satisfied; otherwise, it is the inverse of the strength of the selector.
- The DM between an event and a complex is the product of DMs of individual selectors times the relative t weight of the complex (the ratio of the t weight to the total number of past events).
- The DM between an event and a cover is the probabilistic sum[3] of DMs for complexes in the cover.

The choice of this particular interpretation was experimental. Technical details on the matching function are described by Michalski et al. (1986). A selection of results is shown in Table 4.1, which presents results for three cases of cover reduction:

- "no" (no cover reduction), when the BCR included all complexes that were needed to represent all known cases of the given disease, that is, the *complete* description
- "unique > 1," when the cover included only complexes with a u weight greater than 1
- "best cpx," when the BCR was reduced to the single complex with the highest t weight (the "heaviest")

The system's performance was evaluated by counting the percentage of the correct diagnoses, defined as the diagnosis that receives the highest degree of match and that is considered correct by an expert (see Table 4.1, "Accuracy 1st choice"). For comparison, the table columns "Human experts" and "Random choice" show the estimated performance of human experts (general practitioner/specialist) and the performance representing random choice.

As shown in Table 4.1, the best performance (82%) was obtained, surprisingly, when the BCR consisted of only *one conjunctive statement* ("best cpx") per concept. This representation was also, of course, the simplest, as it required approximately one-fourth the memory of the complete description.

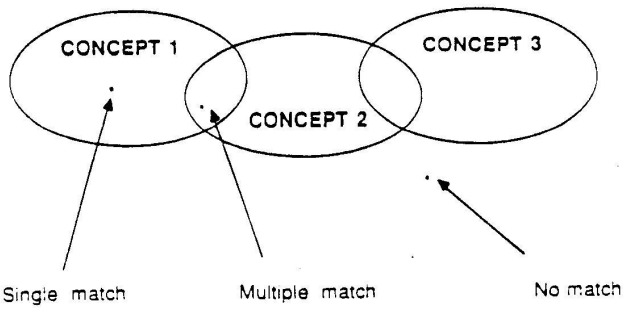


Figure 4.7. Three possible outcomes of matching an event with the base concept representation of different concepts.

The so-truncated cover was then used to diagnose the same new cases, and the performance score was calculated again. The above process was repeated until the truncated cover of each disease had only one complex (termed the *best complex*). Each such iteration represents a different split between the BCR and the ICI. Thus this experiment enabled us to compare the performance of concept descriptions for different distribution between BCR and ICI.

The diagnosis of any new disease case was determined by a simple inferential matching, called *flexible matching*, of the case with the set of covers representing different diseases (here, types of lymphography). This matching treated covers not as logical expressions that are either satisfied or not satisfied by a given case but as descriptions with flexible context-dependent boundaries. The confidence in the diagnosis was defined by the maximum degree of match found between the given case and a cover. Thus the diagnostic decision is determined in the context of all diagnoses under consideration.

The computation of the degree of match distinguished among three possible outcomes of matching an event (here, a disease case) with a set of covers: *single match* – only one cover is strictly matched (i.e., the case completely satisfies only one cover); *no match* – the case satisfies no cover; and *multiple match* – the case satisfies several covers. These three possible outcomes are illustrated in Figure 4.7.

When there is a single match, the diagnosis is defined by the cover satisfied. When there is no match or a multiple match, the degree of (approximate) match is computed for each cover. This computation takes into consideration the strength of conditions represented by individual selectors[2] and t weights of complexes (the t weights are treated as estimates of prior probabilities). The evidence provided by

for storing concept descriptions while preserving the diagnostic performance.

NOTES

The author thanks Doug Medin, Andrew Ortony, Gail Thornburg, and Maria Zemankova for comments and useful criticisms of the earlier versions of this chapter. Remarks of Intelligent Systems Group members P. Haddawy, L. Iwanska, B. Katz, and H. Ko were also helpful.

This work was supported in part by the National Science Foundation under Grant DCR 84–06801, in part by the Defense Advanced Research Projects Agency under grants N00014–85–K–0875 and N00014–87–K–0874 administered by the Office of Naval Research, and in part by the Office of Naval Research under grants N00014–88–K–0226 and N00014–88–K–0397.

1 The above-mentioned analogical match is not to be confused with the analogical mapping discussed in the structure-mapping theory of analogy by Gentner (1983; this volume). The analogical match is related to what Gentner and Landers (1985) call "analogical access." It involves finding semantic correspondences between attributes and relations of the entity to be recognized and the base knowledge representation.

2 The strength of a selector $\{A = R\}$, where R is a set of values of A, is defined as the ratio of the number of all possible values of attribute A over the number of values in R.

3 The probabilistic sum is defined as $p_1 + p_2 - (p_1 \times p_2)$.

REFERENCES

Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles: A problem of polysemy. *Cognitive Psychology, 7*, 167–180.

Barsalou, L. W., & Medin, D. L. (1986). Concepts: Static definitions or concept-dependent representations. *Cahiers de Psychologie Cognitive, 6*, 187–202.

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

Collins, A. M., & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.), *Cognition learning and memory* (pp. 117–137). New York: Wiley.

Collins, A., & Michalski, R. S. (1986). *Logic of plausible inference.* Reports of the Intelligent Systems Group, no. ISG 85–17, Urbana-Champaign: University of Illinois, Department of Computer Science. An extended version appears in *Cognitive Science, 13* (1989).

DeJong, G. (1986). An approach to learning from observation. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach*, (Vol. 2, pp. 571–590). Los Altos, CA: Morgan Kaufmann.

These results show that by using a very simple concept representation (here, a single conjunction) and only a somewhat more complex concept interpretation (as compared to the one that would strictly match the complete concept description) one may significantly reduce the amount of storage required, without affecting the performance accuracy of the concept description. Further details and more results from this experiment are described in Michalski et al. (1986).

This research is at an early stage, and further work is required, both theoretical and experimental. There is, in particular, a need to determine whether similar results can be obtained in other domains of application. Among interesting topics for further research are development and experimentation with more advanced methods for inferential matching and base concept representations, new techniques for representing contexts, and algorithms for learning two-tiered concept representations. For representing physical objects, one needs to develop methods for defining and/or learning permissible transformations of the base concept representations of these objects (e.g., transformations of a typical table that will not change it to some other object). The latter topic is of special importance for understanding sensory perception.

Conclusion

The two-tiered concept representation postulates that the total concept meaning is distributed between a base concept representation and an inferential concept interpretation. The BCR covers the typical, easily explainable concept meaning and may contain a store of examples and/or known facts about the concept. The ICI is a vehicle for using concepts flexibly and for adapting their meanings to different contexts. The inferential interpretation involves contextual information and relevant background knowledge. It may require all types of inference, from truth-preserving deductive inference through approximate deductive and analogical inference to falsity-preserving inductive inference. When dealing with physical objects, the interpretation may involve various transformations of the BCR (e.g., a prototype or a set of prototypes).

Experiments testing some of the ideas on a simple medical example showed that distributing concept meaning more toward inferential concept interpretation than toward the base concept representation (as compared with storing the complete BCR) was highly advantageous, leading to a significant reduction in the size of memory needed

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Sowa, J. F. (1984). *Conceptual structures: Information processing in mind and machine.* Reading, MA: Addison-Wesley.

Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology, 18,* 158–194.

Wittgenstein, L. (1953). *Philosophical Investigations.* New York: MacMillian.

Zadeh, L. A. (1976). A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. *International Journal of Man-Machine Studies, 8,* 249–291.

DeJong, G., & Mooney, R., (1986). Explanation-based learning: An alternative view. *Machine Learning, 1*, 145–176.

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

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

Hofstadter, D. R. (1985). *Metamagical themas: Questing for the essence of mind and pattern*. New York: Basic Books.

Hong, J., Mozetic, I., & Michalski, R. S. (1986). *AQ15: Incremental learning of attribute-based descriptions from examples, the method, and user's guide*. Intelligent Systems Group Rep. no. ISG86–5. Urbana-Champaign: University of Illinois, Department of Computer Science.

Lenat, D. B., Hayes-Roth, F., & Klahr, P. (1979). Cognitive economy in artificial intelligence systems. *Proceedings of International Joint Conference on Artificial Intelligence, 6*, Tokyo.

Lenat, D., Prakash, M., & Shepherd, M. (1986). CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks. *AI Magazine, 6*, 65–85.

Michalski, R. S. (1983). Theory and methodology of inductive learning. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 83–130). Los Altos, CA: Kaufmann.

Michalski, R. S. (1987). Concept learning. In S. C. Shapiro (Ed.), *Encyclopedia of artificial intelligence* (pp. 185–194). New York: Wiley.

Michalski, R. S. (1988). *Dynamic recognition: An outline of the theory*. Report of Machine Learning and Inference Laboratory. Fairfax: George Mason University, Department of Computer Science.

Michalski, R. S., & Chilausky, R. L. (1980). Learning by being told and learning from examples: An experimental comparison of two methods of knowledge acquisition in the context of developing an expert system for soybean disease diagnosis. *International Journal of Policy Analysis and Information Systems, 4*, 125–161.

Michalski, R. S., Mozetic, I., Hong, J., & Lavrac, N. (1986). The multi-purpose incremental learning system AQ15 and its testing application to three medical domains. *Proceedings of the American Association for Artificial Intelligence Conference*, Philadelphia.

Michalski, R. S., & Stepp, R. E. (1983). Learning from observation: Conceptual clustering. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 331–363). Los Altos: CA: Kaufmann.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York: McGraw-Hill.

Minsky, M. (1980). K-lines: A theory of memory. *Cognitive Science, 4*, 117–133.

Mitchell, T. M., Keller, R. M., & Kedar-Cabelli, S. T. (1986). Explanation-based generalization: A unifying view. *Machine Learning, 1*, 47–80.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.

Rosch, E. H. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573–605.