BEYOND PROTOTYPES AND FRAMES: THE TWO-TIERED CONCEPT REPRESENTATION

by

R. S. Michalski

6

Beyond Prototypes and Frames: The Two-tiered Concept Representation*

RYSZARD S. MICHALSKI

Center for Artificial Intelligence, George Mason University, Fairfax, VA 22030–4444, USA

1 Introduction

Cognitive scientists have been, for years, searching for essential ingredients of intelligence. Although this issue may not be satisfactorily resolved for quite some time, two abilities are clearly central to intelligent behaviour. One is the ability to acquire knowledge or skill through experience; that is, the ability to learn. The second is the ability to apply the knowledge or skill possessed to solve new problems; that is, the ability to reason. The new problems may concern actual events in the real world: for example, when one has to react to a new external stimulus; or may be imaginary, for instance, when one creates them for planning purposes.

A precondition for the above abilities is the capability to represent diverse forms of knowledge. As our knowledge is built of individual concepts, to represent knowledge one needs to represent concepts. Consequently, understanding how concepts are represented is a fundamental problem underlying all efforts in the quest to understand intelligence. Although cognitive scientists, psychologists, linguists, philosophers, and artificial intelligence researchers have given considerable attention to this problem, no conclusive solution of it exists.

Other chapters of this book have reviewed some of the most widely known approaches to concept representation. These approaches include the classical view, the probabilistic view, the prototype view, and the frame view – see, e.g., Barsalou and Hale (Chapter 6), Hampton (Chapter 4), Murphy (Chapter 7). Although these views explain many aspects of

concept representation, they do not provide a totally satisfactory and conclusive solution.

The aim of this chapter is to explore a new approach that overcomes some of the weaknesses of these well-known representations. The new approach, called the *two-tiered (TT) concept representation*, considers human concepts not as fixed, well-defined structures, but as flexible, context-modifiable, and background-knowledge-dependent units of our knowledge. To represent concepts so understood, the TT view recommends splitting the total concept representation into two components ('tiers'): one that explicitly defines the most stable and common concept properties, and the second that implicitly defines allowable concept modifications, exceptional cases, and context-dependency. The latter component consists of procedures for 'flexible' matching concept descriptions with instances, and a body of inference rules for characterizing concept variability in different contexts and situations.

As a background for the new approach, the following section outlines our notion of concepts and categories. Later sections provide details about the two-tiered representation, relate it to other views, and finally discuss its application to data analysis.

2 The meaning of concepts and categories

Although everybody agrees that concepts and categories are fundamental building blocks for our knowledge and thinking, there seem to be no universal agreement as to their meaning and definition. There are differences among the views of individual cognitive psychologists on these issues, and differences between the views of cognitive psychologists and researchers on formal theories of learning and artificial intelligence. These issues, however, are important to the proposed two-tiered concept representation, and therefore, we will attempt to clarify our view on them here. We will present a view that appears to us most theoretically satisfying, although it is somewhat different from the views often expressed in cognitive science literature. The main advantages of the proposed view is that it allows us to define more rigorously the meaning of concepts and categories, to distinguish *concepts* from *concept representations*, and to clarify the meaning of concept learning.

Let us start with views expressed recently by some of the most prominent cognitive scientists. Barsalou (1992) states: 'Whereas most theorists would probably reserve "concept" for categorization rules, I will also use "concept" for people's conceptualization of categories.' Medin

(1989) expressed the view that 'a concept is an idea that includes all that is characteristically associated with it' (p. 1469). He also suggests that 'a category is a partitioning or a class to which some assertion or set of assertions might apply' (p. 1469). Smith (1989) views a concept as a 'mental representation of a class or individual and deals with *what* is being represented and *how* that information is typically used during the categorization' (p. 502).

The above characterizations are not all equivalent, and from a viewpoint of a formalist call for some clarification. For example, if concepts are people's conceptualization (or mental representation) of categories, then this implies that there are as many different concepts for the same category as there are different cognitive representations of that category. Everyone has an individual representation of a given category, since people's knowledge about concepts changes, as any new fact adds to the representation. If a concept **square** is the same thing as a cognitive representation of the class of squares, then we have so many concepts of **square** as there are minds (or machines) that hold some representation of **square**.

For similar reasons, one could argue with the view that identifies concepts with a concept description or a classification rule. For any given concept one can usually think of many different concept descriptions or classification rules that specify how to distinguish between concept instances and non-instances. To see this point, suppose that a concept was defined as a certain classification rule, expressed in terms of specific properties and/or relations linked by specific operators. If we were to replace this rule by one syntactically different but logically equivalent (that is, one that produces exactly the same classification), then we would have a contradiction: since the initial and the newly created rule are different from each other, then they would constitute different concepts; yet, since they generate exactly the same classification of instances, they would have to be viewed as the same concept. For example, there can be many different medical procedures (classification rules) for distinguishing a specific type of cancer from other types of cancer. Although these medical procedures are different, the concept of **cancer** is all the same. In short, identifying concepts with their classification rules may lead to a contradiction.

Another concern is that the above views seem to disagree with the view of concepts often expressed in artificial intelligence and formal theories of concept learning (e.g., Fulk and Case, 1990; Sowa, 1984). This view is that concepts are certain sets of entities (the entities are called *concept instances*). A concept *description* is an expression that characterizes this set. Thus, a human or machine can learn a concept, if it can determine a representation that can be effectively used to recognize

concept instances. Thus, a concept is not the same as its representation. The formal status of concepts as sets seems also to be confirmed by our linguistic expressions that use them. It is common to say, 'This example belongs to a concept', or 'The instances of the concept are . . .'. Such expressions seem to be more consistent with the usage of 'concept' as a 'set' than as an 'idea', 'conceptualization', and so forth. The view of concepts as sets has not been common in cognitive science, and therefore may cause difficulties for cognitive psychologists. It allows one, however, to define more rigorously the meaning of concept representation and the similarity between concepts. *Knowing* a concept means to have a *representation* of it that allows one to distinguish concept instances from non-instances.

We will elaborate this view by stating that *concepts are sets of entities grouped together for some reason*. The set constituting a concept may consist of just one individual (e.g., the sun in our solar system; to be sure, since the sun is a dynamic system changing in time, strictly speaking it can be viewed also as a set of instances corresponding to different time moments), a countable number of individuals (e.g., the set of students taking Cognitive Science 101, or the set of birds called blue jay), or an uncountable number of individuals (e.g., the concept of **real number**). There can be many different reasons for grouping entities to a concept. For example, a reason may be that entities have the same function or purpose, have similar appearance, have similar physical, spatial, or temporal characteristics, share the same abstract property (or properties), have the same relation to other entities, or because of some combination of these and/or other factors. Typically, concepts are sets with a name, and are referred to through a name. There can be, however, concepts without a name. For example, entities such as crowds, tanks, speeches, white-red-and-blue flag, and broken monuments of formerly admired communist leaders may constitute an unnamed concept of **'things that come to mind when thinking about the August 1991 revolution in the (former) USSR'**. Because different people may associate different objects with this revolution, this concept is a *personal construct*.

The instances of a concept represent physical or abstract objects. These objects are perceived (by human or machine) through their properties. These properties can be values of certain attributes, relations among objects components, relations among properties of the components, and so on. From now on, attributes or relations used to characterize a set will be called *descriptors*. Thus, every concept instance can be viewed theoretically as a point in a *concept description space* spanned over the descriptors through which the objects are perceived. Concept descriptions then map entities into sets in certain description spaces. In the simplest and the most common case, a concept description space is spanned over

attributes only; that means, zero- or one-argument functions. In a more general sense, a description space can be spanned over any type of descriptors.

When viewing concepts as subsets of a description space, one can characterize a concept instance as more or less typical, depending on the relationship between the instance and other instances. For example, one simple measure of typicality would be the 'distance' between an instance to the 'centre' of the set in the description space. The similarity between concepts can be defined by the 'closeness' between corresponding sets in the concept description space. This definition implies that the similarity between concepts is crucially dependent on what attributes/variables are spanning the concept description space. Without defining the description space, the similarity between two concepts cannot be determined. A description space is often defined only implicitly; for example, via the properties that humans measure through their senses. For example, an orange and a tangerine are usually viewed as similar fruits. The implied assumption is that they are similar in terms of the properties measured by our unaided senses. However, one could think of a description space in which these two entities would be very dissimilar, for example, in the space spanned over specially selected atomic level attributes. In view of these difficulties, the presented 'two-tiered' approach develops a new view of concept typicality.

To recognize an instance of a concept is to identify an a priori known concept (set) to which the given instance belongs. To do so, one always needs to use some knowledge (explicit or implicit) about the concept and about the instance. The knowledge about the concept is usually expressed in the form of a 'concept description' or a 'classification rule'. Such a description or rule is supposed to capture the reason underlying the concept.

The difficulty in understanding the meaning of a concept is that the reason for creating a concept may be incorporated in a body of knowledge that is encoded in the structure of an agent's (concept bearer's) mind as a result of some learning or creative process. Such a body of knowledge may, or may not, be precisely translatable into an effective classification rule. It may be externalized only by the classifications one makes, for example, one may not be able to tell why a given painting is beautiful, but may be quite confident of this evaluation. In sum, the reason (or principle) underlying a concept is not necessarily expressible linguistically, and a concept (a set) should be distinguished from a classification rule for that concept. One basic problem then is: what is the relationship between a concept and a classification (or categorization) rule used for its recognition? Another problem is: what is the relationship between concepts and categories? This section proposes a new view on these

issues, somewhat different from the one currently used in cognitive science literature.

A classification rule is an expression in some language (natural or formal) that constitutes an effective procedure for telling which instances belong to a concept and which do not. Such a procedure uses various attributes and/or relations (that is, descriptors) that are measurable directly or indirectly. As mentioned earlier, there may be many different classification rules for recognizing instances of a given concept. For many concepts, however, only approximate classification rules can be constructed. As mentioned above, this is because the reason for grouping entities to a concept may involve abstract notions expressed in terms of other abstract notions, and these, in turn, may depend on the individual experience of the agent holding the concept.

Consider, for example, the concept of **friendship** or **revolution**. The meaning of such concepts changes with an agent's experience, and thus these concepts are living personal constructs. They are open-ended sets of behaviours that this agent would classify as **friendship** or **revolution**. Such concepts would usually be manifested by instances the agent may generate rather than by any formal descriptions. To define a precise classification rule for such concepts, even only for a specific agent and for a specific time period, one would have to inspect all the different concept instances that the agent could possibly generate. But this may be impossible because the agent may not be able to classify some instances, and/or because the number of instances may be too large. Consequently, some human concepts cannot, in principle, be defined precisely. Consequently, there cannot be precise classification rules for them. For such concepts there can only be approximate classification rules derived from abstract linguistic or other descriptions, or inductively learned from subsets of instances generated by a reasoning agent.

Concepts that were created by defining an a priori classification rule (e.g., prime numbers), or for which an effective classification rule has been developed (e.g., mammals) are called *categories*. Categories are thus special kinds of concepts. An effective classification procedure for a category must be expressed in terms that are either directly observable or measurable, or themselves have an effective recognition procedure. From this viewpoint, concepts such as **peace** or **freedom** are not categories, because no effective recognition procedures for them have been developed. On the other hand, concepts such as '**students who received A in my class**' or '**GMU employees earning less than 35k in 1992**' are categories.

Some concepts are sets that are relatively 'stable', that is, they are more or less permanently useful in describing our world, or our actions in it. Therefore, we develop effective classification rules for them, and they become categories. For example, animal species, plants, diseases,

astronomical objects, and so on, are such concepts. We call them *natural categories*. Other concepts represent sets that may have only temporary utility, like, for example, the concept of '**things to carry out of a burning house**' (Barsalou, 1987), or '**pins in my drawer**'. Such concepts are not natural categories, although they are or can be categories. As stated earlier, for a given concept, there may be many different classification rules. Some of them may be simple but only approximate (e.g., those defining the animal species based on the animal's physical appearance), and some may be complex but precise (e.g., those defining the animal species based on an animal's genetic code). Depending on the need, different types of rules are employed.

Concepts that are sets with *precise boundaries* that do not change with the context or situation in which the concepts are used, are called *crisp* (or *classical*). Crisp concepts occur primarily in science, where – whenever it is useful – they are given a name (e.g., the concepts such as **even number** or **vertebrates**). These concepts remain crisp usually and only if used within the scope of a given scientific discipline; for example, a **triangle** or **rectangle** in geometry, a **group** in mathematics, **water** in chemistry. Outside of this discipline, however, they may represent different sets of entities. For example, a **triangle** may mean an arrangement of streets or cities; a **group** may mean a set of people, and so forth. There are relatively few crisp concepts outside of science. For example, kin relationships, such as **father** or **grandmother**, can be viewed as crisp. But even these concepts are often used in a flexible manner. For example, the founder of a scientific discipline may be called a 'father of that discipline' (e.g., 'Lukasiewicz is a father of multi-valued logic'). Here, the original concept is extended to a larger set through an analogical reasoning.

Concepts that intrinsically do not have precise boundaries, either because they are open-ended and/or because they are context-dependent, are called *flexible*. The boundaries of flexible concepts are imprecise and may dynamically evolve with time, and/or change with the context in which they are used. For example, the concept **computer** (that means, the set of all entities called 'computers') does not have, in principle, a well-defined and context-independent boundary. The entities called computers have been rapidly evolving in time, and it would be unwise a priori to limit the meaning of this concept.

In order to make the above notions clearer, one needs to explain what is meant by *context*. Approximately, by context of a concept we mean a set of concepts *relevant* to the intended meaning of the concept. What concepts are relevant is specified by the agent's background knowledge (sometimes also called the *agent's theory*). A context can be explicitly specified by evoking the name of a more general concept that includes

the concept discussed, or by providing a description characterizing relevant concepts.

In different contexts, a flexible concept may stand for similar or quite different sets of entities. For example, the concept **printer** would have a different meaning in the context of '**old-fashioned office equipment**' than in the context of '**modern office equipment**'. As another example, consider the concept **key**. In the context of **doors, rooms, houses, buildings**, and so on, **key** would mean a set of entities that serve for opening door locks. In the context of **travel, suitcases, briefcases**, and so on, the concept so named would stand for a somewhat different set of entities, related only at an abstract functionality level to the entities in the previous set. Finally, in the context of **problems, solutions, methods**, and so on, the concept **key** would stand for a quite different set, related only by analogy to the previous sets, and could be considered as being a different concept. This example suggests that sets associated with a given name of a flexible concept consist of elements that may be bonded to one another by different relationships, such as identity, surface similarity, analogy, functionality, or just by a certain relationship to other concepts. The context determines which sets are referred to in a given discourse.

Crisp concepts, by the very fact that they have precise boundaries, can be represented by logical-style descriptions that evaluate to 'true' or 'false'. Such descriptions may be logical expressions in disjunctive normal form (DNF), sets of rules, decision trees, grammars, exemplars, prototypes, and so on. These descriptions can always be transformed, by the use of intermediate terms or relations, to a single conjunction of single conditions. Consequently, crisp concepts can be defined by a set of jointly sufficient and individually necessary conditions. Thus, the old controversy associated with the classical view of concept (e.g., Smith, 1989; Smith and Medin, 1981; Sutcliffe, this volume, Chapter 3) can be resolved through the idea of crisp concepts. The classical view simply applies only to crisp concepts. If concepts have precise boundaries, then they can be equally precisely described by a logical description, or by a 'prototype-based' description (but not with the same simplicity). Thus, the prototype view of concepts makes a real difference only if one does not assume that concepts have precise boundaries.

The majority of human concepts are not crisp, but flexible. For such concepts, it may not be possible to construct precise logic-style classification rules. The reason is that boundaries of such concepts – as mentioned earlier – are intrinsically imprecise and/or dependent on the current and future contexts in which the concept may be used. Thus, only approximate classification rules are attainable for flexible concepts.

It should be noted, however, that instances of concepts, crisp or flexible, may occur with different frequencies. One component of the

'typicality' of a concept instance is associated with such a frequency of occurrence. For example, there could be typical or atypical geometrical triangles, in the sense that they occur frequently or infrequently. Since a **triangle** in geometry is a crisp concept (with sharp boundaries), every instance is a 'prototype' triangle. In other words, all geometrical figures satisfying the definition of a geometrical **triangle** are 'equally good' triangles, although they can occur with different frequency. (Another component is 'closeness to BCR'; see below.)

A different situation is with flexible concepts. Since such concepts do not have precisely defined boundaries, or different boundaries apply in different contexts, to develop a classification rule for them one needs to rely on the distributions of instances over a space of the observable descriptors. If there are some instances that belong to the concept frequently and in many situations and many contexts, then they can be viewed as 'highly typical'. Here, typicality can be measured by the degree of similarity (defined in terms of the chosen observable descriptors) between an instance and an 'average' instance in the space spanned over these descriptors. Which descriptors are chosen to characterize an instance depends on how relevant they are to the original principle underlying the concept. A judgement of relevance of different descriptors depends on the knowledge or experience that an agent has with the concept.

The prototype view of concepts (see, e.g., Hampton, this volume, Chapter 4) concerns flexible concepts whose instances have a 'well-behaved' distribution (just one or few modes) in the space spanned over the chosen descriptors. However, for many concepts – for example, **chair, game, house, music, mechanism**, and so on – it is very difficult to determine a set of measurable descriptors that would produce a space in which all their instances would be distributed in such a 'well-behaved' fashion. The key problem in representing such concepts lies in the difficulty of describing all their possible manifestations and context-dependency in terms of measurable properties.

It is a remarkable challenge to understand how people can use and communicate with such concepts without having a priori agreed-upon classification rules or well-defined descriptions. If specifically asked, different people usually produce different descriptions of such concepts. The two-tiered representation proposed below aims at explaining this phenomenon. It can handle both crisp and flexible concepts, but its primary aim is to represent the latter ones.

3 The two-tiered concept representation

The fundamental assumption underlying the two-tiered concept represen-
tation is that total classificatory information about a concept resides not
only in its explicit representation, but also in the inference and matching
procedures applicable to it. Through the application of these inference
and matching procedures the meaning of a concept can be extended,
modified, or adapted to various contexts. The two-tiered representation
thus constitutes a significant departure from the previous approaches,
which attempted to represent the complete meaning of a concept within
one explicit knowledge structure (a list of properties, a logical description,
a set of prototypes). Nevertheless, as explained later, the TT view builds
upon and relates in various ways to other views.

The TT view was originally proposed in Michalski (1986), and
subsequently applied to various problems in concept learning (e.g.,
Bergadano et al., 1988a, 1988b, 1992; Michalski, 1990). The original
formulation of this approach was based on an observation that, although
most human concepts lack precise definition when used outside of a
specific context, they acquire a precise meaning when used in a
combination with other concepts in a specific context. Consider, for
example, the statement: 'This blonde woman near the centre of the room
is a student at GMU'. If there is only one woman with hair of lighter
colour than any other woman near the centre of the room, the statement
above precisely defines the person of interest (although the concepts
blonde, near, and **room** are imprecise by themselves, outside of a specific
context). Thus, statements made of imprecise concepts can convey precise
meaning. In other words, the fact that individual concepts are flexible
(that means, imprecise and context-dependent) does not prevent us from
communicating precise meaning.

The above example illustrates the basic supposition for the TT approach
(Michalski, 1986), that the imprecision of human concepts stems not from
an undesirable vagueness of our concept definitions, but rather from the
universal need for cognitive economy. By allowing individual concepts to
be imprecise and context-modifiable, the expressive power of concepts is
greatly enhanced. The latter means that one can employ fewer concepts
for expressing a greater set of meanings. This leads to a simplification of
descriptions of our immensely complex universe. Various experiments
reported here and elsewhere (e.g., Bergadano et al., 1988a, 1988b, 1992;
Michalski, 1990) have confirmed this idea in a microworld to which it
was applied.

The TT approach assumes that (flexible) concepts have a certain central
tendency and basic usage, which should be described explicitly, as the
'first approximation' of the concept meaning. The complete concept

meaning, however, may vary significantly beyond the basic (or typical) usage, especially in different contexts. The TT view postulates that this variability and context-dependency is best represented implicitly – that means, as results from an application of 'flexible' matching methods and context- and background knowledge-dependent rules of inference. Accordingly, a TT concept representation consists of two components (tiers):

I   The BCR, the *base concept representation*, that represents the basic concept meaning explicitly. This basic meaning may be characterized by specifying the typical function of the concept instances, their usual physical properties and appearance, their structure and the components, their role in a system of concepts by indicating ancestor and descendant concepts, and so forth. (In the specific method used in the experiments mentioned later, the BCR was expressed by a logical DNF description, or an equivalent set of rules.)

II   The ICI, the *inferential concept interpretation*, that consists of inference rules and matching procedures, which modify the basic meaning accordingly to any given situation or context. In matching an instance with the concept representation, these inference rules and matching procedures involve the agent's background knowledge (BK) and the context of discourse. The ICI may involve any type of inference – deductive, analogical, or inductive.

Thus, the basic or typical concept meaning is expressed by the BCR, and imprecision, exceptions, and the context-dependency are handled by ICI. The ICI captures the allowable transformations applicable to a concept in different contexts. Entities that satisfy the BCR are called typical, and those that do not are characterized by a degree of typicality. Factors affecting the degree of typicality are the type and the amount of transformation needed for matching the entity with the BCR, and the frequency of the occurrence of an entity. Highly atypical entities are called *exceptions*, and require ICI rules for identification (Bergadano *et al.*, 1992).

Figure 6.1 illustrates the relationship between the BCR and the ICI in a TT concept representation. It shows that the ICI can, in general, extend the concept meaning beyond the BCR in one area of the description space, and reduce the meaning in another area.

Note that BCR is not a 'prototype', but a characteristic description of a class of entities that are viewed as typical members of a given concept. Such a description may be a disjunction of conjunctions (that means, a DNF expression). A 'prototype' instance should satisfy the BCR, but there could be a number of distinct 'prototype' instances that also satisfy it. For example, a large numer of chairs can be viewed as typical of the concept of **chair**. Also, a chair, a desk, or a bookshelf can be viewed as
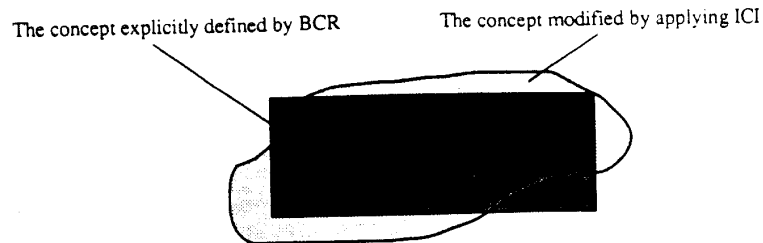
The concept explicitly defined by BCR   The concept modified by applying ICI

Figure 6.1 An illustration of the relationship between the Base Concept Represen-
tation (BCR) and the Inferential Concept Interpretation (ICI)

typical examples of the concept **office furniture**. The BCR can be
expressed at different levels of abstraction. For example, a BCR of the
concept **table** can include statements stating their typical function, typical
physical properties, the typical material they are made of, and so on. In
different situations different components of BCR can be used for concept
recognition. A method of 'dynamic recognition' that uses various subsets
of properties for recognizing a concept in different situations (a situation
is defined by the set of properties that have been observed or can be
observed) have been described in Michalski (1989).

Here is an example of a TT representation of the concept **chair**:

BCR: *Function:* To seat one person.
   Structure: A seat supported by legs and has an attached
   backrest from one side.
   Physical properties: Made of wood or plastic. The seat is
   flat and is located about 14–18 inches above the ground.
   There are usually four legs which are located in the vertices
   of an imaginary square, etc.
   (BCR may also include pictures or 3D models of typical
   chairs.)
   Immediate ancestor (parent) concept: Furniture
   Immediate descendants (children): Straight chair, armchair,
   chaise longue, folding chair, high chair, lawn chair, rocking
   chair, barber chair, throne, wheelchair.

ICI: *Possible variations of the properties in BCR:* The number
   of legs can vary from one to four. The legs may be replaced
   by any support. The shape and the material of the seat, the
   legs and the backrest are irrelevant, as long as the function
   is preserved. The backrest may be very small or missing;
   the seat may be closer to the ground or much higher above

the ground than specified in BCR, and so on.
Situation or context-dependent variations:
Situation: If no typical chair is available → any object on which a person can sit can be called chair.
Context = museum exhibit → an object matching the visual characteristics of a chair is not used for seating persons (any more).
Context = {toys, doll's house, dolls, children playing, . . .} → the size can be much smaller than stated in BCR. The chair does not serve for seating persons, but correspondingly small dolls.
Context = organizational structure, committee, positions of responsibility, University department, etc. → chair means a person being in charge.

To recognize a concept instance, the instance is matched against the BCR of candidate concepts. This process first attempts to find a BCR that directly matches the instance. A match is direct if the BCR's conditions are precisely satisfied by the instance. If there is no direct match, or there are several direct matches, then the process employs the ICI rules of inference and matching procedures. If there is no direct match with any BCR, the system searches for a deductive match, that means, tries to determine if the instance is a logical consequence of some BCR. For example, if there is a BCR condition 'X is a very good student' and ICI has a rule that 'a student who has a high grade average and gets along well with most of the peers and teachers then that student is a very good student', then an instance 'John has high grade average and gets along with almost everyone' deductively matches the BCR condition. If there is no deductive match or there is more than one direct match, a plausible match procedure is employed. Such a procedure determines the degree to which an instance matches different concept candidates by using matching procedures and rules contained in the ICI. The best matching description defines the concept.

Most of the known concept representations, such as semantic network, decision tree, and frame representation, usually employ a direct match. Plausibility matching employs usually a method for determining some degree of similarity or partial match. In a general case, a plausibility matching may employ any type of plausible inference (e.g., approximate deduction, analogy or induction; Collins and Michalski, 1989). Illustrative examples of plausible matching are given in the next section. Performing plausible inferences may involve concept meta-knowledge; for example, the importance of concepts and their frequencies of occurrence, the

relation to other concepts, and other relevant domain knowledge. Such meta-knowledge is assumed to be a part of the second tier.

As mentioned earlier, the ICI includes the background knowledge relevant to the proper interpretation of the concept. Since such background knowledge (called by some authors 'theory', see, e.g., Murphy, this volume, Chapter 7) grows with a person's experience, the same concept may have a different meaning for different people. For example, the concept of a computer is quite different for a novice and a computer expert, who witnessed the evolution of computers over many years. Thus, the two-tiered view provides a simple model for explaining the observable personal differences in concept meaning. In sum, by dividing the concept representation into an explicit and implicit part, the TT view combines elements of the classical representation, prototypical representation, and other elements not present in other representations.

In the cognitively oriented form of the TT representation described here, the distribution of the concept meaning between BCR and ICI is relatively stable. The BCR expresses the general unifying idea, the typical function, and/or measurable properties implied by or correlated with this idea or function. Such a BCR can be viewed as representing the 'first approximation of the concept'. The ICI, in this case, defines the matching procedures and inference rules for handling less typical instances and context-dependency. In a general approach, the TT theory allows any 'distribution' of meaning between BCR and ICI (Michalski, 1990).

An advantage of distributing the concept meaning between the BCR and the ICI is that it permits a learner flexibly to modify or extend the concept meaning by varying matching procedures and inference rules, and/or by changing the context of discourse. The concept meaning can thus be changed without having to alter the BCR. As mentioned earlier, by evaluating the type and the amount of inference involved in matching the BCR with an instance, one may produce a qualitative or quantitative estimation of the degree of typicality.

The ability to produce a measure of the degree of match indicates one principal difference between this approach and the fuzzy set approach (e.g., Zadeh, 1965, 1976). In the fuzzy set approach, a set defining a concept is associated with a membership function, which needs to be *defined* to the learner by a person. The influence of the context is hidden in the definition of this membership function. In the proposed approach, a concept is associated with interpretation procedures and context-dependent rules, which implicitly define the membership of an instance in a concept. These rules and procedures can be used to compute the membership function in different contexts.

4 Examples justifying the two-tiered approach

There are many examples that show that human concepts are not characterized solely by explicitly stated properties, relations, or prototypes. A full description of most concepts involves general and domain-specific rules and background knowledge, in order to capture the concepts' variations, and context-dependency. These rules tell us what transformations of instances preserve or not their concept membership, which are potentially feasible, and what transformations move them outside of a concept. For example, repainting a chair preserves its membership in the concept of **chair**, but shredding it does not. Consequently, to learn a concept one needs to learn not only the BCR, but also the ICI. Let us consider a few examples illustrating and confirming the ideas of the two-tiered representation.

4.1 Example 1: A pizza or a coin?

This example is based on recent experiments by Rips (1989), which involved concepts that can be characterized, using the TT terminology, as having significantly different ICIs, and somewhat comparable values of one attribute in their BCRs. Specifically, a subject was presented with a partial description of an object that expressed only the value of one of its attributes (e.g., its diameter). Based on this description, the subject was supposed to classify the object into one of the two categories. The attribute-value given in the description was somewhere between the subject's average values of this attribute for the two categories. For example, an object was described as 'three inches in diameter', and the categories of choice were **pizzas** and **quarters** (25 cents). The results of experiments were that the subjects consistently classified such an object as likely to be a pizza, although the specified size was closer to the quarter than to a typical pizza.

The two-tiered model explains this behaviour simply. The subjects do not only hold the description of the typical values of objects in each class (the BCR), but also have an understanding of the possible variability of the objects in each category (the ICI). This understanding comes either from a generalization of their observations, or from their relevant background knowledge, for example, about the process that produces objects of the two categories. Consequently, they could admit an object to the category of pizzas, although unusually small for a pizza, but not to the category of quarters, since quarters do not have a variable diameter.

4.2 Example 2: 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 red maple, that slender twigs turn brown, and that the buds are brown and sharp-pointed. As a tree, of course, a maple has a root, trunk, and branches.

Suppose that while strolling on a nice winter day someone tells us that a particular tree is a sugar maple. A simple introspection tells us that the fact that the tree does not have leaves would not strike us as a contradiction of what we know about sugar maples. Yet, clearly, the presence of leaves of a particular shape is deeply embedded in our typical image of a maple tree. The two-tiered theory explains this phenomenon simply: the inferential concept interpretation associated with the general concept of deciduous trees evokes a rule 'in winter deciduous trees lose leaves'. Since a maple is a deciduous tree, the rule would apply to the maple tree. The result of this inference would override the stored standard information about maple trees, and the inconsistency would be resolved. Matching an instance with the concept requires in this case deductive reasoning from the knowledge associated with a more general concept.

4.3 Example 3: A coffee pot used as a bird feeder

This example is taken from a study of children concepts performed by Keil (1989). In this study, children were shown objects representing some known concepts. These objects then underwent a sequence of transformations that produced an object resembling an instance of a different concept. In one experiment, the transformations were realistic, that means, corresponded to changes that could be obtained by some simple operations. In the other experiments, the transformations were purely imaginary (they did not correspond to changes that would be allowed by the ICI).

For example, in one case the initial object was a **coffee pot**. Subsequently, the pot was punched full of holes, and its spout was removed. It was then filled with bird feed and hung from a branch of a tree. In another case, the initial object (presented as a picture) was a **raccoon**. It was then presented as having been transformed to black with a white tail, and able to produce a super smelly stuff from beneath its tail, like a skunk. The children were then asked if the end result of these transformations was, in the first case, a **bird feeder** or a **coffee pot**, or in the second case, a **raccoon** or a **skunk**. The experiments thus can be interpreted as testing

the children's understanding of the feasibility of various transformations of instances of different concepts.

The results of the first experiment were that all children (five to nine years old) indicated that it was a coffee pot that was turned into a bird feeder. In the second experiment, however, all children experienced a difficulty in accepting the change of the raccoon into a skunk. These experiments indicate that children knew not only physical characteristics of the objects in the involved categories (here, the **coffee pots** and **raccoons**), but also had the knowledge of transformations that are applicable to objects in these categories (i.e., the ICI).

4.4 Example 4: The concept of a triangle

Let us consider the concept of **triangle**. In geometry, a triangle is a plane figure consisting of three non-co-linear points connected by straight lines. Using the notation of *annotated predicate calculus* (APC), which is like predicate calculus, but employs more compact forms of some logical expressions (Michalski, 1983), one can write:

$$\text{Triangle}(T, P_1, P_2, P_3)$$

$\leftarrow$ Consists$(T, P_1 \& P_2 \& P_3) \&$
Type$(P_1 \& P_2 \& P_3, \text{point}) \&$
Connected_by_straight-line$(P2 \& P1, P3 \& P2, P1 \& P3) \&$
Non-co-linear$(P1, P2, P3)$

In the above expression, symbol '&' is used in two related meanings: first, to denote an ordinary (*external*) conjunction connecting predicates; and second, to denote an *internal conjunction* – that means, a conjunction of terms, which is treated as a *compound* argument of a predicate. For example, the predicate 'Type(P1 & P2 & P3, point)' states that P1 and P2 and P3 are points, thus is equivalent to a conjunction of three predicates: Type(P1, point) & Type(P2, point) & Type(P3, point).

Suppose that someone tells us that the tall towers in his/her hometown form a big triangle. Obviously, the meaning of the triangle in this statement differs from that in the formal geometrical description. To match the statement with the concept of triangle, the following assumptions and transformations need to be made:

I. In the context of describing a configuration of physical objects such as towers, the individual objects play the role of points. Thus, the statement implies that towers correspond to points, and that there are three towers in the town. The matching operation involves drawing an analogy between the abstract points and the towers, which can be characterized as consisting of one step of generalization (GEN):

Point —GEN→ Object

and one step of specialization (SPEC):

Object —SPEC→ Tower

II. In the context of towers, the presence of a 'straight line' is imaginary – that means, there is no physical connection – but one could imagine a straight line between the objects (towers). The condition 'Connected_by' is then satisfied in such an abstract sense. This is an operation of abstraction. Thus, matching the statement about a triangular arrangement of towers with the definition of a triangle involves here a generalization, specialization, and abstraction. (A reader interested in an analysis of the differences between generalization and abstraction may consult Michalski, 1991.)

The examples above show that relating a concept instance to a concept representation is not just a straightforward comparison of attribute-values in an instance with those in the concept representation, as done in various mechanized decision processes. They show that such a process may involve different forms of inference, as postulated in the two-tiered representation.

5 Learning two-tiered concept representations

The essence of the two-tiered representation is a recognition that a concept may have explicit and implicit properties. The explicit properties are represented by logical-style expressions, and the implicit ones by a set of rules and/or matching procedures. Consequently, to acquire a TT concept representation, one needs to learn both the explicit (BCR) and the implicit (ICI) components of a TT representation.

In general, each concept has a different BCR that has to be learned. On the other hand, the ICI is often the same or very similar for various classes of concepts. Consequently, it can be learned only once for a given class of concepts. All concepts of this class, and/or descendent concepts in the type hierarchy will share the same ICI. For example, all diseases in the class of liver diseases may share the same flexible method for matching the BCR descriptions with symptoms. By sharing the ICI among concepts, a significant economy of the concept representation can be achieved.

The two-tiered approach does not impose restrictions on what type of language is used to represent BCR and ICI parts of a concept. Any type of knowledge representation formalism could potentially be employed. It

is assumed that the language employed should depend on the goal of the concept representation. The goal implies the degree of abstraction with which the concept needs to be represented, and this sets the criteria for the representation language. For example, in a diagnostic situation, it may be sufficient to represent a disease at a high level of abstraction, like by an attributional description in the form of propositional calculus rules. In a medical research situation, the representation may have to be much more complex, and involve not only structural representations (e.g., using a predicate calculus language), but also images and drawings.

In most of the experiments done so far with the TT representation, the representation language was an attributional calculus, called VL_1. In this calculus, descriptions are expressed in the form of decision rules. A decision rule links a logical product of elementary conditions with a decision (e.g., a class membership). The elementary conditions relate a multiple-valued attribute to a value or a set of values. For example, elementary conditions may be [sex = male], or [colour = blue or red] or [weight 2 kg]. (See, e.g., Bergadano *et al.*, 1992; Michalski, 1974, for details on this representation.) In experiments in learning structural descriptions, an extension of predicate calculus (called *annotated predicate calculus*) was used (Bergadano *et al.*, 1988a).

The rest of this section outlines an implemented method for learning two-tiered concept representations from concept examples using the above mentioned attributional calculus. It also presents results from comparing the method with several other methods that use different concept representations. Full details of the method and the experiments performed are described in Bergadano *et al.* 1988a, 1988b; 1992).

In the method, the *Base Concept Representation* (BCR) is created in two phases. In Phase I, a complete and consistent concept description is induced from the supplied examples. This is done by an inductive learning programme AQ-15 (Michalski *et al.*, 1986). Such a description explains (covers) all positive examples of the concept being learned, and none of its counter-examples (negative examples). The description is in the form of rules that link single conjunctive conditions with a concept name.

In Phase II, the obtained description is optimized according to a *description quality criterion*. The criterion prefers the descriptions that are computationally simple and at the same time have a high predictive accuracy; that means, correctly classify new concept instances. The optimization process involves making various changes in the original description, such as removing a rule, removing a condition, or modifying a condition in a rule. These changes correspond to either generalization or specialization of the original rules.

For example, rule removal is a specialization operation, and condition removal is a generalization operation. A condition modification can be

either a specialization or generalization operation depending on the type of modification. The rules obtained as a result of the optimization process may be significantly simpler than the original rules.

The obtained rules, however, may not directly match all the positive examples, that means, they may constitute an *incomplete* concept description, and/or may directly match some negative examples (that means, may constitute an *inconsistent* concept description). It turns out, however, that the incompleteness and/or inconsistency is often well compensated by the inferential concept interpretation (ICI), so that the resulting rules may actually have a higher predictive accuracy than the original rules.

Depending on the type of description learned, the ICI may consist of just a procedure for *flexible matching*, or the procedure plus a set of rules that capture the concept exceptions. In the latter case, the rules are acquired through an interaction with a teacher.

The method has been implemented in POSEIDON (AQ16), and experimentally tested on two real-world problems: learning the concept of an acceptable union contract, and learning voting patterns of Republicans and Democrats in the US Congress. For comparison, a few other learning methods were also applied to the same problems. These methods included simple variants of exemplar-based learning, an ID3-type decision tree learning (ASSISTANT programme; Chestnik, Kononenko, and Bratko, 1987), a method that utilizes complete and consistent descriptions, and a method that uses a very simple BCR (the 'Top rule' description) and flexible matching.

To give the reader some sense of the results obtained, Table 6.1 presents a summary of the results from testing these methods (based on Bergadano *et al.*, 1988a, 1988b, 1992).

The 'simple exemplar-based method' illustrates the exemplar view of concepts; that means, represents concepts by sets of their examples. An unidentified concept instance is recognized by determining examples that are most 'similar' or 'close' to it. In the 'k-nearest' procedure, k most similar examples are determined, and the concept associated with the majority of them is assigned to the instance (in the experiments reported, k was 1, 3 and 5).

The predictive accuracy of each description expresses the percent of classification of testing examples that are evaluated as correct by an expert. The complexity of the description was measured by the number of rules in it, and the total number of elementary conditions in all the rules.

The 'complete and consistent' concept description is a disjunctive normal expression (equivalent to a set of rules) that describes all positive examples, and none of the negative examples. The 'top rule' description

Table 6.1 Summary of the results for testing descriptions learned by different methods

| | Problem | | | |
| | Labour contract | | Congress voting | |
Method	Accuracy (%)	Complexity[a]	Accuracy (%)	Complexity[a]
Simple exemplar-based				
1-nearest neighbour	77	27/432	86	51/969
3-nearest neighbour	83	27/432	84	51/969
5-nearest neighbour	80	27/432	84	51/969
Complete and consistent description (AQ15 without rule truncation)	80	11/29	86	10/32
'Top rule' description (AQ15 with rule truncation)	83	2/6	85	2/6
Optimized two-tiered description (AQ16 or POSEIDON with rule truncation and condition optimization)	90	9/12	92	10/21

[a] The complexity of each description is expressed as a pair (number of rules/number of conditions).

represents a very simple variant of two-tiered representation. The BCR is just a single rule that covers the largest number of examples among all the rules in the consistent and complete description. Such a rule can be viewed as representing the most typical concept examples. The ICI is a procedure for *flexible matching* that determines a similarity between an instance and the BCR of candidate concepts. The concept associated with the most similar BCR is assigned to the unidentified instance. It is interesting to notice that, although this form of two-tiered representation is extremely simple, it still has a relatively high predictive accuracy.

The 'optimized two-tiered' description is a more advanced variant of the TT representation. The BCR is a set of rules generated by removing some rules and/or conditions from the consistent and complete description. This is done by the so-called TRUNC/SG optimization procedure (Bergadano et al., 1988a, 1988b, 1992). The procedure generalizes some parts and specializes other parts of the consistent and complete description in order to maximize a *general description quality criterion*. The ICI of the concept was a set of expert-generated rules for covering exceptional cases identified by the system. As shown in Table 6.1, the 'optimized

TT' description had the highest predictive accuracy, and was also the next simplest, after the 'top rule' description.

6 The tiered view versus other views

As mentioned earlier, cognitive science literature distinguishes among several views of concept representation. The major ones are the classical view, the prototype (or probabilistic) view, the exemplar view, and the theory view (Allen, Brooks, and Norman, 1988; Medin and Smith, 1984; Murphy and Medin, 1985; Nosofsky, 1988; Smith and Medin, 1981; and in this volume: Barsalou and Hale – Chapter 5; Hampton – Chapter 4; Murphy – Chapter 7). Some authors also distinguish between different *types of formalisms* (or representation languages) employed in the concept representation; for example, feature lists, decision trees, connectionist models, predicate logic expressions, simple and complex frames (e.g., Barsalou and Hale, this volume, Chapter 5). The distinction between a 'view' and a 'representational formalism' is usually not well explained in the literature. Our interpretation is that a view of a concept concerns inherent aspects of concepts as sets, while the representational formalism is a language in which those aspects are expressed. In principle, a sufficiently rich representational formalism – for example, predicate calculus – with a proper interpretation, could be used for expressing any of the concept views.

The classical view assumes that concepts have well-defined borders, and can be represented by singly necessary and jointly sufficient conditions. This implies that they can be represented by logic-style conjunctive definitions. The inadequacy of this view has been widely recognized (e.g., Barsalou and Medin, 1986; Lakoff, 1987; McCloskey and Glucksberg, 1978; and Wittgenstein, 1922). To overcome the limitations of this view, other views have been advanced, such as the *prototype view* and the *exemplar view* (e.g., Allen, Brooks, and Norman, 1988; Medin and Smith, 1984; Nosofsky, 1988; Smith and Medin, 1981).

The prototype view represents concepts by prototypes, and uses the so-called *family resemblance* principle (e.g., Rosch and Mervis, 1975). Unlike the classical view, this view does not require that all concept instances share the same properties. It can, therefore, explain the observed phenomenon that instances of the same concept may differ in their *typicality*. The exemplar view claims that concepts can be adequately represented by their exemplars (e.g., Bareiss, Porter, and Craig, 1990; Smith and Medin, 1981). Both these views can also be criticized on

various grounds. The prototype view, which formally is based on the idea of linear separability, disregards the existence of correlations between the attributes, the context-dependency, and other information that has been shown to be relevant to human concept understanding (e.g., Estes, 1986; Flannagan, Fried, and Holyoak, 1986).

The exemplar view employs similarity-based and context-sensitive matching, which has received support in the cognitive science literature. It is also supported by an observation that people do store examples of concepts, especially exceptional examples. This view disregards, however, the importance of general concept descriptions that clearly play a role in human concept representations. Such general descriptions are useful in many ways; for example, for quickly identifying the differences between concepts, for recognizing them from partial information, for handling context-dependency, for efficiently storing invariant information about concepts, and so on. The above operations would be difficult to perform if concepts were represented only by examples. In some work using the exemplar view, general aspects of concepts are captured under the idea of 'category structure', which is a network of domain knowledge that specifies the relevance of exemplars to the concept they define (Bareiss, Porter and Craig, 1990).

Some recent work has advocated a theory view (or knowledge-based view) which emphasizes the need to define concepts through their role in theories in which they exist as interrelated components (Carey, 1985; Hofstadter, 1985; Murphy, this volume, Chapter 7; Murphy and Medin, 1985; Schank, Collins, and Hunter, 1986).

The two-tiered (TT) view constitutes a departure from the existing approaches, although it has a close relationship to most of them. It assumes that concepts have a certain central tendency, and proposes to describe this tendency explicitly, in the form of a logic-style description. Such a description is the 'first approximation' of the concept. On the other hand, it also assumes that concepts may have a great variability and be context-dependent, and that these aspects are best handled not explicitly, but implicitly, by a flexible matching method and context- and background knowledge-dependent rules of inference.

Since the TT view proposes to represent basic concept properties (BCR) in the form of explicit logic-style expressions, it is related to the classical view (it admits, however, that the BCR can be in the form of conjunctive descriptions, as well as disjunctive descriptions). It also recognizes that most concepts have a central tendency, and therefore one needs to distinguish between more or less typical concept examples. In this sense, the TT view is similar to the probabilistic view and the fuzzy set concept representation (e.g., Zadeh, 1965, 1976). Unlike these views, it proposes to handle this aspect by the inferential concept interpretation

(ICI) that employs a flexible concept matching procedure (Bergadano *et al.*, 1992).

The TT view has also a relationship to the exemplar view, as it postulates the use of sophisticated matching procedures and inference rules in classifying new instances. Also, the TT view recognizes the usefulness of storing individual examples, especially in the case of exceptions (e.g., Zhang, 1990a, 1990b). The TT approach is also closely related to the theory view, as it stresses the role of background knowledge and inference in concept representation, especially in the case of non-typical or borderline instances. The main differences between the TT view and the theory view are that the former makes a distinction between the explicit and implicit representation, and postulates that the agent's knowledge or 'theory' plays a role primarily in the inferential concept representation, rather than in the basic concept representation.

7 Summary and open problems

The most significant aspect of the presented two-tiered view is that it represents concepts as a combination of an explicit and implicit representations. This stands in contrast to other views that attempt to compact concept representation into a single monolithic structure, either as a conjunction of properties (as in the classical model), or as a linear weighted function (as in the prototype model). In the two-tiered representation, the first tier, the base concept representation (BCR), captures the explicit and common concept meaning, and the second tier, the inferential concept interpretation (ICI), defines allowable modifications of the base meaning and exceptions. The typical concept instances strictly match the BCR, and thus can be recognized efficiently. The exceptional and/or context-dependent cases involve the ICI, which takes more time, but is done relatively less often. Such a two-tiered representation is particularly suitable for learning flexible concepts; that means, concepts that inherently lack a precise definition and are context-dependent. The TT view asserts that it is impossible, in principle, to give such concepts a precise meaning in an explicit and context-independent form. Therefore, it proposes to describe precisely only their central tendency, and to use flexible matching procedures and inference rules to characterize the concept's variability, to resolve borderline cases, and to express their context-dependency.

The TT view seems to be confirmed by an observation that people can usually recognize rapidly and without any difficulty typical concept

instances; for example, that chair is an example of **furniture**. When, however, they need to decide about special borderline cases (for instance, if a lamp or painting is an example of **furniture**), they try to resolve the problem by reasoning and/or an evaluation of the similarity to the cases with clear membership. This seems to contradict approaches that try to resolve such cases by weighting features or statistical evaluation. In the TT view, the reasoning and similarity evaluation is conducted by matching procedures, inference rules and agent's background knowledge residing in the second tier of the concept representation.

The TT view also assumes that the imprecision of human concepts does not stem from an undesirable vagueness of our concept definitions, but rather from the universal need for cognitive economy. By allowing concepts to have a context-modifiable meaning, and making them precise only to the extent to which a given situation and/or context requires them to be precise, the expressive power of concepts is greatly enhanced. This means that one can employ fewer concepts for expressing more meanings, and this helps us to simplify our descriptions of our immensely complex universe. Thus, by compacting the variability of concepts into rules and matching procedures, rather than trying to express it explicitly, one can reduce the overall complexity and improve the effectiveness of concept representation. The experiments on learning TT representations have demonstrated that the obtained TT representations tend to be considerably simpler and have better predictive accuracy than other representations.

The ideas of TT representation are recent, and many problems remain unsolved. Among especially interesting and important problems is how to represent not just one, but a complex system of concepts using the TT representation. As indicated in this chapter, the ICI usually can be shared or inherited from higher-level concepts. This problem is then how to allocate different parts of the ICI within a structure of TT representation. There is also a need for a clearer analysis of the relationship between the TT view and other views, and experimental studies with human subjects to determine the adequacy of the TT model as a cognitively viable concept representation. There are also many problems related to the development of systems capable of learning TT representations, and efficiently using them for concept recognition processes. In particular, an important problem is how people learn individual tiers of different concepts.

Finally, a theoretical analysis and more experimental investigations are needed to confirm or disconfirm the initial finding that the TT representation can lead to significant memory savings and a simultaneous improvement of the predictive accuracy of concept representations over other representations.

Note

References

Allen, S. W., Brooks, L. R., and Norman, G. R. (1988) 'Effect of prior examples on rule-based diagnostic performance', *Proceedings of the Twenty-ninth Annual Meeting of the Psychological Society* (Nov.).

Bareiss, E. R., Porter, B. W., and Craig, C. W. (1990) 'Protos: an exemplar-based learning apprentice', in Y. Kodratoff and R. S. Michalski (eds), *Machine Learning: an Artificial Intelligence Approach* (vol. 3), Palo Alto, CA: Tioga, Morgan Kaufmann.

Barsalou, L. W. (1987) 'The instability of graded structure in concepts', in U. Neisser (ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization* (pp. 101–40), Cambridge University Press.

Barsalou, L. W. (1992) *Cognitive Psychology: an Overview for Cognitive Scientists*, Hillsdale: Erlbaum.

Barsalou, L. W. and Medin, D. L. (1986) 'Concepts: fixed definitions or dynamic context-dependent representations', *Cahiers de Psychologie Cognitive*, 6: 187–202.

Bergadano, F., Matwin, S., Michalski, R. S., and Zhang, J. (1988a) *Learning Flexible Concept Descriptions Using a Two-tiered Knowledge Representation: Ideas and a Method* (Report 88–4), Fairfax, VA: George Mason University, Artificial Intelligence Center, Machine Learning and Inference Laboratory.

Bergadano, F., Matwin, S., Michalski, R. S., and Zhang, J. (1988b) *Learning Flexible Concept Descriptions Using a Two-tiered Knowledge Representation: Implementation and Experiments* (Report 88–5), Fairfax, VA: George Mason University, Artificial Intelligence Center, Machine Learning and Inference Laboratory.

Bergadano, F., Matwin, S., Michalski R. S., and Zhang, J. (1992) 'Learning two-tiered descriptions of flexible concepts: the POSEIDON system', *Machine Learning Journal*, 8: 5–43 (Report 91–3, Fairfax, VA: George Mason University, Artificial Intelligence Center, Machine Learning and Inference Laboratory).

Carey, S. (1985) *Conceptual Change in Childhood*, Cambridge, MA: MIT Press.

Chestnik, B., Kononenko, I., and Bratko, I. (1987) 'ASSISTANT 86: a knowledge elicitation tool for sophisticated users', in I. Bratko and N. Lavrac (eds), *Proceedings of the Second European Working Session on Learning*, Bled, Yugoslavia (May).

Collins, A., and Michalski, R. S. (1989) 'The logic of plausible reasoning: a core theory', *Cognitive Science*, 13: 1–49.

Estes, W. K. (1986) 'Memory storage and retrieval processes in category learning', *Journal of Experimental Psychology: General*, 115: 155–75.

Flannagan, M. J., Fried, L. S., and Holyoak, K. J. (1986) 'Distributional expectations and the induction of category structure', *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12: 241–56.

Fulk, M. A. and Case, J. (1990) *Proceedings of the Third Annual Workshop on Computational Learning Theory (COLT 90)*, Rochester, NY, University of Rochester (Aug.).

Hofstadter, D. R. (1985) *Metamagical Themas: Questing for the Essence of Mind and Pattern*, New York: Basic Books.

Keil, F. (1989) *Concepts, Word Meanings, and Cognitive Development*, Cambridge, MA: Harvard University Press.

Lakoff, G. (1987) *Women, Fire and Dangerous Things: What Categories Tell Us about the Nature of Thought*, Chicago: Chicago University Press.

McCloskey, M. and Glucksberg, S. (1978) 'Natural categories: well-defined or fuzzy sets?' *Memory and Cognition*, 6: 462–72.

Medin, D. (1989) 'Concepts and conceptual structure', *American Psychologist*, 44: 1469–81.

Medin, D. L. and Smith, E. E. (1984) 'Concepts and concept formation', *Annual Review of Psychology*, 35: 113–18.

Michalski, R. S. (1974) 'Variable-valued logic: System VL1', *Proceedings of the 1974 International Symposium on Multiple-Valued Logic* (pp. 323–46), Morgantown, WV: West Virginia University (May).

Michalski, R. S. (1983) 'Theory and methodology of machine learning', in R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (eds), *Machine Learning: an Artificial Intelligence Approach*, Palo Alto, CA: Tioga: Morgan Kaufmann.

Michalski, R. S. (1986) 'Concept learning', in S. C. Shapiro (ed.), *Encyclopedia of Artificial Intelligence* (vol. 1, pp. 185–94), New York: Wiley.

Michalski, R. S. (1986) 'Two-tiered concept representation, analogical matching and conceptual cohesiveness', Paper presented at the Workshop on Similarity and Analogy, Allerton House, University of Illinois (June). (Extended version in S. Vosniadou and A. Ortony [eds], *Analogy and Similarity* pp. 122–45, New York: Cambridge University Press, 1989.)

Michalski, R. S. (1989) *Dynamic Recognition: an Outline of a Theory on How to Recognize Concepts Without Matching Rules* (Report), George Mason University, Center for Artificial Intelligence, Machine Learning and Inference Laboratory.

Michalski R. S. (1990) 'Learning flexible concepts: fundamental ideas and a method based on two-tiered representation', in Y. Kodratoff and R. S. Michalski (eds), *Machine Learning: an Artificial Intelligence Approach* (vol. 3), San Mateo, CA: Morgan Kaufmann.

Michalski, R. S. (1991) 'Toward a unified theory of learning: an outline of basic ideas', *Proceedings of the First World Conference on the Fundamentals of Artificial Intelligence*, Paris (July).

Michalski, R. S.; Mozetic, I., Hong, J., and Lavrac, N. (1986) 'The AQ15 inductive learning system: an overview and experiments', *Proceedings of the American Association for Artificial Intelligence Conference (AAAI)*.

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

Nosofsky, R. M. (1988) 'Examplar-based accounts of relations between classification, recognition and typicality', *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14: 700–8.

Rips, L. J. (1989) 'Similarity, typicality, and categorization', in S. Vosniadou and A. Ortony (eds), *Analogy and Similarity* (pp. 21–59), Cambridge: Cambridge University Press.

Rosch, E. and Mervis, C. B. (1975) 'Family resemblances: studies in the structure of categories', *Cognitive Psychology*, 7: 573–605.

Schank, R. C., Collins, G. C., and Hunter, L. E. (1986) 'Transcending induction category formation in learning', *The Behavioral and Brain Sciences*, 9: 639–86.

Smith, E. E. (1989) 'Concepts and induction', in M. I. Posner (ed.), *Foundation of Cognitive Science* (pp. 501–26), Cambridge, MA: MIT Press.

Smith, E. E. and 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.

Wittgenstein, L. (1922) *Tractatus Logico-philosophicus*, London: Routledge & Kegan Paul.

Zadeh, L. A. (1965) 'Fuzzy sets', *Information and Control*, 8: 338–53.

Zadeh, L. A. (1976) 'A fuzzy-algorithmic approach to the definition of complex or imprecise concepts', *International Journal of Man-Machine Studies*, 8: 249–91.

Zhang, J. (1990a) 'Learning flexible concepts from examples: employing the ideas of two-tiered concept representation', Unpublished doctoral dissertation, Department of Computer Science, University of Illinois at Urbana-Champaign.

Zhang, J. (1990b) 'A method that combines inductive learning with exemplar-based learning', *Proceedings of Tools for Artificial Intelligence*, IEEE Computer Society Press, Washington, DC.