1

UNDERSTANDING THE NATURE OF LEARNING:

Issues and Research Directions

Ryszard S. Michalski*
Massachusetts Institute of Technology

Abstract

This chapter presents an overview of goals and directions in machine learning research and serves as a conceptual road map to other chapters. It investigates intrinsic aspects of the learning process, classifies current lines of research, and presents the author's view of the relationship among *learning paradigms, strategies*, and *orientations*.

1.1 DO WE NEED LEARNING MACHINES?

Artificial intelligence (AI) is now experiencing extraordinary growth, and applications of its ideas and methods are appearing in many fields. Among its most visible and important successes are the development of expert systems, practical implementations of natural language-understanding systems, significant advances in computer vision and speech understanding, and new insights into building powerful inference systems. This rapid expansion of activities in AI leads one to believe that new successes are forthcoming.

---

*On leave of absence from the University of Illinois at Urbana-Champaign.

In this context, it is important to ask what the limitations of the current methods are and what new directions research in this field should take. One of the obvious limitations, and hence a direction for further research, relates to *machine learning*—a field concerned with developing computational theories of learning and constructing learning systems.

Except for experimental programs developed in the course of machine learning research, current AI systems have very limited learning abilities or none at all. All of their knowledge must be programmed into them. When they contain an error, they cannot correct it on their own; they will repeat it endlessly, no matter how many times the procedure is executed. They can neither improve gradually with experience nor learn domain knowledge by experimentation. They cannot automatically generate their algorithms, formulate new abstractions, or develop new solutions by drawing analogies to old ones, or through discovery. Generally speaking, these systems lack the ability to draw *inductive inferences* from information given to them. One might say that almost all current AI systems are *deductive*, as they are able to draw conclusions from knowledge incorporated and/or supplied to them, but they cannot acquire or generate new knowledge on their own.

By contrast, when we look at human intelligence we see that among its most striking aspects are the abilities to acquire new knowledge, to learn new skills, and to improve with practice. In time, use of these learning abilities can turn a young, inexperienced person into a journeyman engineer, educator, artist, or physician. Our common perception is that a person who would repeat the same error again and again could hardly be called intelligent. The ability to learn from error is considered fundamental to the individual and to the society at large (Popper, 1959, 1981; Kuhn, 1970; Lakatos, 1970; Berkson and Wettersten, 1984; Hayes-Roth, 1983—*Machine Learning I*, chap. 8, see below).

Because learning ability is so intimately entwined with intelligent behavior and research in AI gives us new insights and powerful tools to study it, many researchers postulate that one of the new central goals for research in artificial intelligence should be understanding the nature of learning and implementing learning capabilities in machines (McCarthy, 1983; Schank, 1983). Overcoming the above-mentioned limitations sets an agenda of research tasks.

Questions then arise about whether this goal is achievable, and if so, whether it is desirable. Let us start with the question of achievability. Answering it involves us immediately in questions of definition. Can we identify some general criteria such that, if satisfied by a machine, we would agree to call this machine a learning system?

As machine learning research has shown, learning ability manifests itself not as an all-or-nothing quality but as a spectrum of information-processing activities, ranging from the direct memorization of facts and acquisition of simple skills by imitation to very intricate inferential processes leading to creation of new concepts and discovery of new knowledge. It always involves a change in a system, whether human or machine, that makes it better in some sense.

For now, let us put the question of the definition of learning aside (it is discussed in more detail in the next section of this chapter) and observe that machine learning is experiencing a renaissance after its past steady but slow growth. Efforts to develop programs exhibiting some forms of learning have multiplied in recent years. This young field has already achieved a number of successes. A summary of some of these efforts is found in *Machine Learning: An Artificial Intelligence Approach* (Michalski, Carbonell, and Mitchell, 1983), henceforth referred to as *Machine Learning I*. The current book is a sequel; it reports some key subsequent efforts characteristic of the state-of-the-art in machine learning.

On the basis of the results achieved so far, it is clear that some rudimentary machine learning abilities are possible. Already there exist programs able to formulate new concepts and discover previously unknown regularities in data; develop decision rules that can outperform human rules; draw interesting analogies; automatically learn problem-solving heuristics; or develop generalized plans for achieving a goal. Many of these programs are discussed in *Machine Learning I*. What is less clear is the level of progress that can be achieved in machine learning using conventional computer hardware and present programming methods. As always in science, such questions can be answered only by conducting further research and continuing to develop experimental learning systems.

New dimensions of research in machine learning will open with the development of *connection machines*, *fifth generation computer systems*, and other novel computer architectures, currently under development (e.g., Hillis, 1981; Kawanobe, 1984). For example, Hinton, Sejnowski, and Ackley (1984) describe how learning may occur in *Boltzmann machines*. The knowledge acquired by such systems is represented by the strengths of the connections between simple, neuron-like elements. The research in this direction should address the problem of overcoming the limitations of early systems of this type, such as the *Perceptron* (Minsky and Papert, 1969). New potential for research in machine learning also emerges in connection with the development of new programming systems, in particular, logic programming and its first embodiment in PROLOG (Robinson, 1983).

Why is it desirable to develop learning machines? It appears that the development of such systems is necessary to ensure further progress in artificial intelligence or closely related disciplines. This seems to be particularly true in areas such as expert systems or any large-scale, knowledge-based systems; computer vision and speech understanding; natural language understanding; intelligent tutoring systems; and (truly) friendly human-machine interfaces. As more and more complex tasks are set for AI systems, more and more knowledge must be represented in them. Such knowledge must encompass domain-specific facts and rules, commonsense heuristics and constraints, and general concepts and theories about the world. The scope of knowledge in any system must be widened to avoid a common problem with the current systems, sometimes referred to as falling off the *knowledge cliff* (Feigenbaum, 1984) or *brittleness* (Holland, 1975, chap. 20; see also Larkin et al., 1985). The problem is that a system performs well within the scope of knowledge provided to it, but any slight move outside its narrow competence causes the performance to deteriorate rapidly.

Introducing all the required knowledge into any new system is a very complex, time-consuming, and error-prone process, requiring special expertise. For example, building an expert system involves a collaborative effort of highly trained experts—at least one *domain expert* and one *knowledge engineer* (Davis and Lenat, 1982; Hayes-Roth, Waterman, and Lenat, 1983; Buchanan and Shortliffe, 1984). This task can be simplified by using machine learning techniques. Such techniques would enable a system to develop decision rules from examples of experts' decisions and through the automated analysis of facts in a database.

With the rapid increase in the amount of data and knowledge that the society generates, there is a growing need not only for storing, organizing, and delivering this information but also for using it in new, creative ways. Knowledge can be viewed as *compressed information* (Rendell, 1983), and we now need machines that can compress databases and information systems into knowledge bases automatically via conceptual analysis of their contents. As envisioned by Michie (1982), "*the most technically gripping challenge, even if not immediately the most economically important, will be how to spread the computer wave from the front end of the scientific process, the telescopes, microscopes, . . . spark chambers and the like, back to the recognition and reasoning processes by which the chaos of data is finally consolidated into orderly discovery.*"

This chapter's author might add that the computer will have a role not only as *scientists*' and *technologists*' intelligent assistant but also as an intelligent *personal* assistant. Individuals in the expanding information society will need such assistants to cope with the overwhelming amounts of available information and the complexities of everyday decision making. In order for such assistants to play the designated roles, their function and knowledge should by *dynamic*. These assistants should be able to adapt to changing demands and be self-modifiable; that is, they should be able to learn.

A similar need for learning abilities exists in the areas of computer vision and speech understanding. To build a computer vision system, one has to incorporate into it a variety of vision-specific transformations; concepts of geometry; physical and functional descriptions of visual objects the system is to recognize; and related information (e.g., Winston et al., 1983; Winston, 1984). To "handcraft" all this information into a system is difficult. It would be much easier to teach the system by showing it examples of given concepts and have it learn the appropriate generalizations and descriptions, just as we teach visual concepts to humans.

A system capable of understanding and interacting with humans in natural language has to be equipped with knowledge of syntactic properties of language (Marcus, 1980), as well as with many concepts and concept structures (such as frames, scripts, and schemata) capturing semantic and pragmatic aspects of the language (Winograd, 1981; Schank, 1982; see also chaps. 19 and 21 of this volume). One may estimate that in an advanced natural language understander, the number of such concepts and concept structures may easily reach tens of thousands or more. Programming all this knowledge into a computer is a monumental task. It is very desirable to simplify this task by employing a learning system. In addition, even if at some point all this knowledge were incorporated in a machine, a language understander would not work well for long without learning abilities. The meaning of human concepts is dynamic; it changes with time and adapts to new contexts and requirements. Novel concepts are continuously being created and developed, and some are being outgrown. Therefore, as in the cases above, we need a learning system capable of acquiring new concepts and concept structures by generalization from examples or by analogy to prior knowledge. Such a system should be able to modify, specialize, or generalize old concepts in a flexible fashion.

Intelligent tutoring systems must be able to present material at a level of difficulty and detail suited to the state of knowledge of the student. In order to do so, the system must know and follow the student's changing knowledge. A desirable way of acquiring this information is not by repeated direct testing but by learning from clues, behavior, and the implicit model of the student during tutorial sessions. Thus learning abilities are required not only from the student but from the tutor as well (Sleeman and Brown, 1982; Sleeman, 1983—*Machine Learning I*, chap. 16).

Through learning capabilities future computers should be able to acquire knowledge directly by using documents and books, by conversing with humans, and by generalizing observations of their environment, which they make with their artificial senses. They should be capable of improving through practice and experience. It is possible that future machine learning systems will suffer little, if at all, from some human limitations, such as poor memory, distracted attention, low efficiency, and the difficulty of transferring acquired knowledge from one learner to another. Once one learning system is developed, a theoretically limitless number of copies of it can be built, which, one hopes, can be employed to learn new knowledge in diverse domains. In addition, any new knowledge acquired by a learning system can be copied to other systems rapidly and relatively inexpensively (unlike human knowledge, which must be painstakingly reacquired by each new student).

Of course, we are still far away from such idealized vision, but it has now become conceivable that such learning systems might be developed in the future. It is then desirable to consider not only expected advantages but also possible undesirable consequences. The latter issue could be dismissed by observing that any new technology brings new opportunities for misuse, and that this has never stopped us from developing it. Moreover, such aspects are usually considered an issue outside scientific or technical research. Yet we need to examine this particular issue carefully, for the creation of machines that can self-acquire knowledge brings about new dimensions of complexity in the development of technology and reflects on the way the field of machine learning should be developed.

The first dimension of complexity is the predictive opacity of self-changing systems. Predicting the behavior of machines that can learn inductively is considerably more difficult than predicting the behavior of machines without such an ability. The key idea behind learning machines is that they should be able to create knowledge that can surprise their human creators. This might cause unexpected difficulties, or even dangers, if someone would apply such a system to solve important problems without understanding the system's limitations. In addition, the increased unpredictability of learning machines implies increased possibilities for their misuse.

Some experts argue that predicting behavior of complex computer systems is very difficult already. They look at the addition of learning capabilities to our computers as further amplification of these difficulties, but not as a quantum leap to a new state. Whether we see a leap or merely an amplification of unpredictability, a strong expectation is that potential benefits from this technology will amply compensate for such undesirable consequences. And with regard to the increased potential for its misuse, why not use these smart learning machines to "police" other machines to prevent or combat attempted misapplication?

In addition to the difficulty of predicting the behavior of learning machines, another dimension must be considered, which stems from the very nature of any knowledge other than factual observation. As has been observed by Hume (see, e.g., 1888) and later by Popper (1959) and others, such knowledge is *inherently conjectural;* that is, any knowledge created by generalization from specific observations or by analogy to known facts cannot in principle be proven correct, though it may be disproven.

This results from the fact that inductive inference is not *truth preserving* but only *falsity preserving* (Michalski, 1983). As an illustration, consider this statement: "All scientists at MIT's AI Laboratory are bright." A deductive conclusion from this statement can be that Roger Light, who works at the AI Laboratory, must be bright. If the original premise is true, then this conclusion must be true also. An example of inductive inference from the initial premise might be this statement: "All scientists at MIT are bright." In this case, even if the original premise were true, such an inductive conclusion might not be. However, if the original premise is false, then this inductive conclusion must be false also. Thus, in contrast with a deductive system, correct inputs to an inductive system do not guarantee the correctness of the outputs. Moreover, for any given input there is theoretically an infinite number of possible inductive conclusions. The ones we actually make reflect the preferences, assumptions, and constraints that we use in formulating our generalizations (Medin, Wattenmaker, and Michalski, 1985; Utgoff, chap. 5).

For the above reasons, if learning machines are to generate knowledge useful to us, it is important that they be equipped with knowledge of all the relevant human constraints and assumptions. As it is unlikely that *all* subtle human and societal constraints and preferences will ever be made known to machines, there is the possibility that machine-generated knowledge will violate some human constraints. A quote from Hofstadter (1980) is pertinent here: "*Unless* [the program] *had an amazingly faithful replica of human body . . . it would probably have enormously different perspectives on what is important, what was interesting, etc.*" Because the perception of what is important and what is interesting is a necessary component in guiding creation of new knowledge (Lenat, 1983), such differences are significant. Thus when such machine-created knowledge is used, it may lead to solutions that are technically flawless but socially undesirable.

A related concern is that people may give too much credibility to the knowledge created by machines. This phenomenon has already been observed in related contexts, for example, when people are unduly influenced by results of computer statistical analysis without clearly understanding its assumptions, or when people ascribe personality to a computer consultation system, as in the case of ELIZA (Weizenbaum, 1976). Furthermore, even if it may be well known to scientists that inductively generated knowledge is inherently error-prone, this fact may be less obvious to nonexperts.

An important implication of the above discussion is that any new knowledge generated by machines should be subjected to close *human scrutiny* before it is used. This suggests an important goal for research in machine learning: If people have to understand and validate machine-generated knowledge, then machine learning systems should be equipped with adequate *explanation facilities*. Furthermore, knowledge created by machines should be expressed in forms closely corresponding to human descriptions and mental models of this knowledge; that is, such knowledge should satisfy what this author calls the *comprehensibility principle* (Michalski, 1983). When designing explanation capabilities for learning systems, one should strive to facilitate human understanding not only of the surface results but also of the underlying principles, assumptions, and theories that lead to these results.

One may hypothesize that although the existence of advanced learning machines would eliminate the current *knowlege acquisition* bottleneck, it could ultimately create a *knowledge ratification* bottleneck. In this situation so much new knowledge might be generated by machines that it could become difficult for human experts to test and approve it. Should this happen—well, future researchers will have an interesting problem with which to while away their idle hours. One may envision these researchers inventing sophisticated learning machines that would design experiments to test knowledge created by other sophisticated learning machines.

With these notes of concern, mixed with arguments stressing the importance of machine learning, let us now look more closely at the intrinsic properties of the learning process.

1.2 WHAT IS LEARNING?

As mentioned earlier, a common view holds that learning involves making changes in the system that will improve it in some way. In this description, the term *improve* needs more precision. Clearly, wine improves with time, but nobody would call such an improvement learning.[1] Simon (1983—*Machine Learning I*, chap. 2) gives a more precise characterization:

> *"Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more effectively the next time."*

The requirement that a system improve performance for learning to take place is widely accepted. There are, however, activities that can be categorized as learning, in which the *improvement criterion* is difficult to apply (as will be seen in a discussion below). Minsky (1985) in his insightful theory of thinking, *The Society of Mind*, replaces this criterion with a more general one requiring that changes are merely useful:

> *"Learning is making useful changes in our minds."*

He subsequently observes that such a definition is too broad to be of any use. Let us then approach the problem of capturing the fundamental aspects of learning in another way. It may be observed that learning is often equated simply with acquiring *new* knowledge, as in the statement: "As the satellite burned in the atmosphere, the spacelab astronaut *learned* that the satellite had an auxiliary antenna." In this case, the astronaut simply acquired a piece of information, but this will never improve his performance with *this* satellite. The *knowledge acquisition* aspect of learning seems to be the *essence* of most learning acts. Those acts where it appears to play only a small role are cases of what is usually termed *skill acquisition*. The latter refers to gradual improvement of motor or cognitive skills through repeated effort, sometimes involving little or no conscious thought (Carbonell, Michalski and Mitchell, 1983—*Machine Learning I*, chap. 1). In this discussion, however, we will concentrate on the knowledge acquisition aspect of learning, a theme that recurs throughout the book.

In order to acquire knowledge of anything, one, obviously, has to represent this knowledge in some form, whether as declarative statements, procedures, a mixture of the two, or as something else (McCarthy, 1968). This fact and the above considerations lead us to the following characterization of learning:

> *Learning is constructing or modifying representations of what is being experienced.*

---

[1]This counterexample was suggested by Steve Tanimoto from the University of Washington in Seattle.

The concept of *experience* includes here any sensory stimuli, as well as internal *Gedanken* processes. These stimuli and internal processes are the vehicles through which the learning system perceives the reality that it is trying to represent. The internal thought processes can themselves be a subject of learning.

Thus, from the above view, the central aspect of learning is the process of constructing a representation of some reality, rather than improving performance. Performance improvement is considered to be a consequence and often the purpose of building the representation, but it can be asserted only in the context of the learner's goals. Because most learning acts indeed involve improvement of performance and because it is easier to measure performance than to read minds, naturally we link the two. Yet, performance improvement does not seem to be an invariable condition for every act of learning. There are situations in which it does not appear to be of primary relevance, as in learning to appreciate beauty. There are also situations in which it may even be misleading. The latter situations occur when it is difficult to accurately assess the learner's goal. For example, workers in a labor camp may want to learn how to do less and appear to do more, yet they keep this goal secret. From the viewpoint of an external observer, these workers will appear not to be learning, as their performance will be decreasing with practice. Thus it seems clear that to determine learning by measuring performance may not be possible without knowing the goals of the learner.

Three dimensions seem to be particularly important for evaluating the constructed representations: *validity*, *effectiveness*, and *abstraction level*. Validity (or truthfulness) refers to the degree of accuracy with which the representation fits the reality. It characterizes the precision of the mapping between the reality and the representation. The second criterion, effectiveness, attempts to capture the performance aspect of learning. It characterizes the usefulness of the representation for achieving a given purpose or goal. The more effective the representation, the better the performance of the system. Thus this criterion is central for tasks in which performance is of primary concern. The third criterion, abstraction level, reflects the scope, detail, and precision of concepts used in the description. It defines the *explanatory* power of the representation. These three dimensions together determine what may be called the *quality* of learning.

The representations can be in the form of symbolic descriptions, algorithms, simulation models, control procedures, plans, images, or general formal theories. If one stretches the concept of representation to include physical or physiological imprints occurring in the nervous system when one is acquiring a skill, the above view of learning seems also to cover skill acquisition.

From this viewpoint, a fundamental problem in any research on machine learning concerns the form and method used to represent and modify the knowledge or the skill being acquired. With regard to the question of modifying knowledge, it is important to identify the components and the properties of the representation that are modifiable by the system and those that are not.

In the taxonomy of machine learning research given in chapter 1 (Carbonell, Michalski, and Mitchell) of *Machine Learning I*, three criteria were suggested as especially useful for classifying and comparing machine learning investigations: *learning strategy, knowledge representation*, and *application domain*. The learning strategy refers to the type of inference employed by the system during learning. Some additional ideas reflecting recent progress on this topic are presented in section 1.4 below. The criteria of knowledge representation and application domain were well covered in the above-mentioned reference and will be omitted here. Instead, two other classification criteria will be discussed in some detail: *research paradigms* (section 1.3) and *learning orientations* (section 1.5). The research paradigm criterion refers here to the approach taken to construct a system, and the learning orientation refers to the scope and the subject of study.

1.3 RESEARCH PARADIGMS

Since the inception of machine learning in the fifties, research efforts have placed the emphasis at different times on different approaches and goals. One can distinguish three major research paradigms or approaches in this area: *neural modeling* and *decision theoretic techniques; symbolic concept acquisition;* and *knowledge-intensive, domain-specific learning*. These research approaches differ chiefly in the amounts of a priori knowledge *built into* the learning system and in the way knowledge is *represented* and *modified* in the system.

The neural modeling approach strives to develop general-purpose learning systems that start with little initial knowledge. Such systems are usually referred to as *neural nets* or *self-organizing* systems. A system of this type consists of a network of interconnected elements, typically neuron-like, that perform some simple logical function, usually a threshold logic function. Such a system learns by incrementally modifying the *connection strengths* between the elements, typically by changing continuous (i.e., non-discrete) weights associated with these connections. The system's initial knowledge is provided by the choice of the input elements that represent selected attributes of objects under consideration and by the structure and initial strength of the connections in the network. This can be a random structure, one prearranged by the designer, or a mixture of the two. Such learning systems include the *Perceptron* (Rosenblatt, 1958), *Pandemonium* (Selfridge, 1959), and any learning machine using *discriminant functions* (Nilsson, 1965). More recent examples stemming from this paradigm are various adaptive control systems (Tsypkin, 1972; Caianiello and Musso, 1984). Research in this area has led to the *decision-theoretic approach* in pattern recognition. Related to this approach is research on *evolutionary learning* (Fogel, Owens, and Walsh, 1966; Conrad, 1983) and on *genetic algorithms* (Holland, 1975; see also chap. 20). As mentioned earlier, there is a resurgence of interest in this learning paradigm with the recent efforts to develop *connection machines* (Hinton, Sejnowski, and Ackley, 1984).

Characteristic features of systems built under this paradigm include low levels of a priori built-in knowledge and the use of continuously changeable parameters to achieve learning. A related feature is the numerical character of learning methods and algorithms. This strongly contrasts this paradigm with the next two paradigms, in which the main emphasis is on creating and manipulating complex symbolic structures during the process of learning.

In *symbolic concept acquisition* (SCA), the system learns by constructing a symbolic representation of a given set of concepts through the analysis of examples and counterexamples of these concepts. The representations typically are in the form of a logical expression, a decision tree, production rules, or a semantic network. Some of the systems developed under this paradigm have multipurpose applicability and have demonstrated practical usefulness. Examples of such systems are Winston's ARCH program (Winston, 1975), the AQVAL program (Michalski, 1975), and ID3 (Quinlan, 1979). In this paradigm, the attributes or predicates relevant to the concept are provided to the system by the teacher.

In *knowledge-intensive, domain-specific learning* (KDL), the system contains numerous predefined concepts, knowledge structures, domain constraints, heuristic rules, and built-in transformations relevant to the specific domain for which the system is built. Not all the relevant attributes or concepts are proved initially; the system is expected to derive new ones in the process of learning (this author refers to such a process as *constructive induction*). Thus the main differences between the KDL and SCA paradigms lie in the amount and the kind of background knowledge supplied to the system and the richness of knowledge structures generated by the system. Learning systems based on this approach are typically developed for a specific domain and cannot be used directly in another domain. The research in this paradigm has explored not only the strategy of *learning from examples*, but also strategies such as *learning by analogy*, and *learning by observation and discovery* (see section 1.4). Examples of systems based on this approach are Meta-DENDRAL (Buchanan and Feigenbaum, 1978) and AM (Lenat, 1983—*Machine Learning I*, chap. 9).

Many systems developed in the past represent a certain mixture of the above-mentioned approaches. An interesting combination of the SCA and KDL approaches is a system based on the idea of an *exchangeable knowledge module*. Such a system combines general-purpose learning mechanisms with built-in facilities for defining and using domain-specific knowledge. When such a system is applied to a given problem, domain-specific knowledge is supplied to it by the teacher via the system's knowledge representation facilities. By separating general inference capabilities from the domain-specific knowledge, such a learning system can be applied to a wide spectrum of different domains and still take advantage of domain-specific knowledge in the process of learning. This philosophy underlies the INDUCE system, which learns structural descriptions of objects from examples (Michalski, 1980). Winston's program for learning by analogy is another example (Winston, 1982). The LEX system for acquiring and refining problem-solving heuristics (Mitchell, Utgoff, and Banerji, 1983—*Machine Learning I*, chap. 6) and the EURISKO program for discovering new heuristics (Lenat, 1983) are other examples. Several chapters in this volume describe learning methods that also fall into this category.

For a historical review of these three research paradigms the reader is referred to chapter 1 in *Machine Learning I*. A sample of contemporary research on self-organizing systems is found in Caianiello and Musso (1984). A recent review of approaches to machine learning has been made by Langley and Carbonell (1984). The primary concerns of this book are symbolic concepts acquisition and knowledge-intensive, domain-specific paradigms of learning.

1.4 LEARNING STRATEGIES

In every learning situation, the learner transforms information provided by a teacher (or environment) into some new form in which it is stored for future use. The nature of this transformation determines the type of learning strategy used. Several basic strategies have been distinguished: *rote learning, learning by instruction, learning by deduction, learning by analogy*, and *learning by induction*. The latter subdivides into *learning from examples* and *learning by observation and discovery*. These strategies are ordered by the increasing complexity of the transformation (inference) from the information initially provided to the knowledge ultimately acquired. Their order thus reflects increasing effort on the part of the student and correspondingly decreasing effort on the part of the teacher. In any act of human learning, a mixture of these strategies is usually involved. It is useful to distinguish these strategies not only for tutorial purposes but for the purpose of designing learning systems as well. Though most current systems focus on a single learning strategy, one may expect that machine learning research will give increasing attention to multistrategy systems. Chapter 1 of *Machine Learning I* describes these learning strategies in detail. Because of their importance to this book and because of some changes in their classification brought about by recent research, they will be reviewed briefly here.

In rote learning there is basically no transformation; the information from the teacher is more or less directly accepted and memorized by the learner. A major concern here is how to index the stored knowledge for future retrieval. In learning by instruction (or *learning by being told*), the basic transformations performed by a learner are *selection* and *reformulation* (mainly at a syntactic level) of information provided by the teacher. In deductive learning, the learner draws deductive, truth-preserving inferences from the knowledge given and stores useful conclusions (this strategy was identified as a separate category only recently; see Michalski, 1983, 1985). Deductive learning includes knowledge reformulation, knowledge compilation, creation of macro-operators, caching, chunking, equivalence-preserving operationalization, and other truth-preserving transformations (see Glossary).

If the transformation process involves generalization of input information and selection of the most plausible or desirable result, that is, the inductive inference, then we have inductive learning. Learning by analogy is deductive and inductive learning combined. Here, descriptions from different domains are matched to determine a common substructure, which serves as the basis for analogical mapping. Finding the common substructure involves inductive inference, whereas performing analogical mapping is a form of deduction. *Learning by being reminded*, described by Schank (1982), can be viewed as a form of learning by analogy. Learning by analogy is discussed in chapters 13 (Burstein), 14 (Carbonell), and 15 (Dershowitz).

Inductive learning can be subdivided into learning from examples and learning by observation and discovery. In *learning from examples* (also called *concept acquisition*), the task is to determine a general description explaining all positive examples and excluding all negative examples of the target concept. The examples are provided by a source of information, which can be a teacher who knows the concept or the environment on which the student performs experiments and from which it receives feedback. The latter case is called *learning by experimentation* (this includes *learning by doing* and *learning by problem solving*). *Stimulus-response* learning can also be classified as a form of learning from examples.

Recent research has revealed two interesting subdivisions within this form of learning: *instance-to-class* and *part-to-whole generalization*. In instance-to-class generalization, the system is given independent instances (examples) of some class of objects, and the goal is to induce a general description of the class. Most research done on learning from examples has concentrated on such instance-to-class generalization. The objects can be structured blocks, geometrical shapes, descriptions of diseases, stories, problem solutions, control operators, and so forth. Various aspects of this problem are discussed in chapters 3 (Winston), 5 (Utgoff), 6 (Quinlan), 7 (Sammut and Banerji), 8 (Lebowitz), and 9 (Kodratoff and Ganascia). For a review of earlier methods for such generalization, see Dietterich and Michalski (1983—*Machine Learning I*, chap. 3) and Cohen and Feigenbaum (1982).

In part-to-whole generalization, the task is to hypothesize a description of a whole object (scene, situation, process), given selected parts of it. For example, given a collection of snapshots of selected parts of a room, reconstruct the total view of that room. Another example is to determine a rule (a theory) characterizing a sequence of objects or a process from seeing only a part of this sequence or process. This type of learning problem is considered in chapter 4 (Dietterich and Michalski). A closely related area of research concerns the *qualitative process prediction* (Michalski, Ko, and Chen, 1985).

In learning by observation and discovery (also called *descriptive generalization*), one searches, without the help of a teacher, for regularities and general rules explaining all or at least most observations. This form of learning includes *conceptual clustering* (forming object classes describable by simple concepts), constructing classifications, fitting equations to data, discovering laws explaining a set of observations, and formulating theories accounting for the behavior of a system. *Genetic algorithms* (Holland, chap. 20) and *empirical prediction algorithms* (Zagoruiko, 1976) can be viewed as variants of this learning strategy. Various aspects of this strategy are discussed in chapters 16 (Langley et al.), 17 (Stepp and Michalski), 18 (Amarel), and 19 (DeJong).

The primary focus of this book is on learning by induction and analogy. Therefore, a few additional comments may be useful about inductive inference, which is at the heart of these strategies. Inductive inference starts with a set of facts (observations)—and optionally with an a priori hypothesis about these facts—and produces a preferred generalization explaining these facts. As mentioned before, it is a falsity-preserving inference accomplished by the application of *generalization inference rules* (Michalski, 1983a). As noted by Popper (1981) and others, "pure" induction, that is, direct inference from facts to theories without any *interpretive* (*explanatory*) concepts, is impossible. These concepts are needed to describe the observations and are part of the learner's *background knowledge.* This background knowledge is a necessary component of any inductive process. It also includes goals of learning, domain-specific constraints, causal relationships, heuristics and biases that guide the generalization process, and the criteria for evaluating competing hypotheses.

One can distinguish two techniques for guiding and constraining generalization: the *similarity-based* and the *constraint-based techniques.* The similarity-based technique explores *inter-example* relationships; that is, it examines the examples and counterexamples of a concept in order to create a concept description. It searches for features shared among facts or examples in the same class and looks for common causes and explanations of why different examples belong to the same class. It generalizes over the differences between examples either by ignoring the differing features or by formulating concepts that encompass the differences. Some early learning methods using this technique are reviewed by Dietterich and Michalski in chapter 3 of *Machine Learning I.*

The constraint-based technique exploits the *intra-example* relationships, which constrain the interpretive or explanatory concepts applied to one or more facts or examples. Any generalization of these facts or examples must satisfy these constraints. For example, when generalizing the fact that a box is on the table, one should satisfy the constraint that whatever is on the table cannot be so heavy that it would break the table or so large that it could not be placed on the table. A variant of this technique is described by Andreae (1984), who uses the concept of *justification* for a hypothesis. Another important variant, called an *explanation-based* generalization, puts the emphasis on the role of explanatory knowledge (Mitchell, Keller, and Kedar-Cabelli, 1986). It applies a system's background knowledge to formulate a high-level conceptual explanation or interpretation of a given fact or event. In chapter 19, DeJong discusses a method implementing such a technique in the context of story understanding. The similarity-based and constraint-based techniques are complementary and can be used simultaneously in learning systems.

1.5 LEARNING ORIENTATIONS

The previous two sections discussed two important classifying criteria for machine learning research: *learning paradigms* and *learning strategies*, respectively. To recapitulate, the first criterion concerns the type of knowledge represented and manipulated in the system, and the second criterion deals with the type of inference performed on the knowledge. This section will briefly discuss one more classifying criterion, the *research orientation*, which concerns the scope and subject of study. By analogy, a *paradigm* corresponds to one's point of departure and the terrain through which one travels, a *strategy* specifies the means of locomotion, and an *orientation* indicates the destination.

As described in chapter 1 of *Machine Learning I*, research in machine learning encompasses three interconnected orientations:

- Theoretical analysis and development of general learning algorithms
- The development of computational models of human learning processes (also called *cognitive modeling*)
- Task-oriented studies concerned with building learning systems for specific applications (also called an *engineering orientation*)

Research in the first orientation investigates theoretical learning tasks, or simplified practical ones, and tries to develop algorithms that accomplish these tasks independently of application. There is no restriction on the type of algorithm developed. The algorithm need not be similar to the one a human might use to perform the given task. As a variation, some authors postulate that at least the knowledge structures generated as an end result of learning should be similar to those a human being might create, although the process of their creation can be different (Michalski, 1983a). In this orientation researchers strive to chart the theoretical space of possible learning algorithms. Chapters 3 (Winston), 5 (Utgoff), 7 (Sammut and Banerji), and 9 (Kodratoff and Ganascia) represent a sample of work representative of this orientation.

In the second orientation, human learning is the focus, and the development of computational theories and experimental models of human learning is the goal. This research will likely have an important influence on human education as well as on the techniques of implementing machine learning systems. Chapters 10 (Rosenbloom and Newell), 11 (Anderson), and 14 (Carbonell) are characteristic of this orientation.

Finally, work in the third orientation undertakes specific practical learning tasks and tries to develop engineering systems capable of performing these tasks. An example here would be a program that learns to recognize dangerous conditions for aircraft in flight. Such efforts usually have to address a host of other problems not directly related to learning, such as the appropriate interpretation of the input signals or the development of problem-specific transformations of the data. Any useful ideas from the other two orientations are readily adopted in this orientation. Often, when a solution to a specific problem is found, it is generalized to a method for solving a class of similar problems. An example of such research is described by Dietterich and Michalski in chapter 4.

The above three research orientations make up a trichotomy of mutually dependent and supportive efforts that fuel the machinery of learning research. Such a trichotomy has come to pervade the whole of artificial intelligence.

1.6 READER'S GUIDE TO THIS BOOK

As indicated in sections 1.3 and 1.4 above, this book is concerned with the SCA (symbolic concept acquisition) and the KDL (knowledge-intensive domain-specific learning) paradigms and concentrates on inductive and analogical learning strategies. Both major types of inductive learning—that is, learning from examples and learning by observation and discovery—are represented. The chapters are grouped into six parts reflecting the major learning strategy or the research orientation employed in the work.

Part One provides an introduction and discussion of general issues in the field of machine learning. After the overview presented in this chapter, views from several researchers on important problems in this field for the decade of the eighties are presented in **chapter 2.** These topics emerged from a panel discussion held at the 2nd International Machine Learning Workshop at the University of Illinois in June 1983 (Michalski, 1983b).

Part Two describes a selection of results on *learning from examples*. In **chapter 3**, Winston integrates ideas about several interrelated topics: learning from precedents and exercises, using *near misses* in learning, generalizing *if-then* rules, and employing *unless* conditions to prevent incorrect rule application. The role of an unless condition is to block a given if-then rule whenever facts at hand satisfy this condition. Such a condition facilitates an incremental improvement of rules.

In **chapter 4**, Dietterich and Michalski present a theoretical framework and methodology for a certain type of *part-to-whole generalization*. They describe a general method using three models for discovering a rule that characterizes a sequence of objects and predicts a plausible sequence continuation. Each object in the sequence is described by discrete attributes, which are either given a priori or derived by applying various inference rules and sequence transformations.

Utgoff in **chapter 5** investigates the role of *bias* or *preference criterion* in determining a plausible hypothesis in inductive learning. He presents a methodology and a program—STABB—for shifting bias in the course of learning from examples.

In **chapter 6,** Quinlan examines the effect of noise in training examples on the discovery of classification rules and their accuracy. He makes several interesting conjectures about how to formulate the learning task when training examples are expected to contain noise.

Next, in **chapter 7,** Sammut and Banerji investigate the role of previously learned concepts in the learning of new ones and the problem of inductive learning with an *active* learner. Such a learner is not just passively accepting examples from a teacher but is also generating examples on its own and asking the teacher whether they represent the concept being learned.

In **chapter 8,** Lebowitz discusses a somewhat related problem. He explores the use of concepts stored in the memory for generalizing complex structural descriptions. His *Generalization-Based Memory* method determines what concepts to learn and formulates definitions of the concepts learned. The ideas are exemplified by two programs, one for concept evaluation, the other for generalization of complex structural descriptions.

Next, in **chapter 9,** Kodratoff and Ganascia discuss various theoretical aspects of the generalization process. They show how generalization is accomplished by creating links among training examples. These links are represented as variable bindings.

Part Three takes up *cognitive aspects of learning.* In **chapter 10,** Rosenbloom and Newell present ideas about modeling processes that underlie improvement of performance by practice. Their model of practice is based on the concept of *chunking,* that is, grouping subgoals into higher goals. They show that this model explains the known *power law* of human practice.

Next, in **chapter 11,** Anderson discusses learning mechanisms involved in *knowledge compilation*, that is, in the process by which subjects move from a declarative representation of a skill to a procedural representation. He shows how mechanisms of *composition* (collapsing multiple productions into a single production) and *proceduralization* (building into productions information that resides in declarative form in the long-term memory) can simulate the initial stages of skill acquisition in the domain of learning how to program.

In **chapter 12,** Forbus and Gentner present their work on a computational model of human learning of physical domains. They use *Qualitative Process Theory* to model human physical knowledge and *Structure Mapping Theory,* which characterizes analogy and other comparisons, to describe processes of changing knowledge representations.

Part Four focuses on the topic of *learning by analogy.* Burstein, in **chapter 13,** presents a model of learning by analogical reasoning. He describes it in the context of acquiring the semantics of assignment statements in the BASIC programming language. According to his model, the use of analogies to learn concepts in a new domain depends strongly on causal abstractions previously formed in a familiar domain. These analogies are extended incrementally to handle related situations.

In **chapter 14,** Carbonell presents his theory of *derivational analogy* and its implications for case-based reasoning and expertise acquisition. In essence, the derivation of solutions to related problems is replayed and modified to solve new and increasingly more complex problems. The method is proposed as a means of automating knowledge and skill acquisition for expert systems.

Dershowitz, in **chapter 15,** focuses on analogy as a tool for automatic programming. He shows how analogies between program specifications (as well as between their derivations) can be used to debug a program or to modify an existing program to perform a new task. These analogies can also be used to derive an abstract schema of a set of programs and to instantiate a schema in order to yield a particular program.

Part Five covers *learning by observation and discovery.* In **chapter 16,** Langley, Zytkow, Simon, and Bradshaw describe four systems addressing different aspects of scientific discovery. BACON.6 formulates empirical laws characterizing any numerical observational data. GLAUBER takes on discovery of qualitative laws of chemical reactions. STAHL undertakes the problem of determining components of substances involved in such reactions. Finally, DALTON focuses on the formulation of structural models for these reactions.

In **chapter 17,** Stepp and Michalski report on their recent work on *conceptual clustering,* that is, creating a classification of observations by identifying subclasses that correspond to simple concepts. Unlike previous work on generating goal-free classifications of unstructured objects, the new research takes on the construction of goal-oriented classifications of structured objects. The authors describe and illustrate by examples how a learner's concepts and inference rules are used in constructing such purposive classifications.

In **chapter 18,** Amarel discusses problems of theory formation in the context of program synthesis. He illustrates his method and ideas by a problem of inferring a program from input-output data associations in the domain of partially ordered structures. His method emphasizes the role of algebraic and geometric models and the importance of shifting problem representations in the program synthesis task.

Taking a different tack, DeJong in **chapter 19** discusses a method of learning from observation that exploits the inner constraints among explanatory concepts in the system's background knowledge to guide the process of generalization from a single example. His examples are stories about people's problem-solving behavior. This knowledge-based generalization process is used to propose new schemata.

Part Six explores some general aspects of learning. In **chapter 20,** Holland discusses general-purpose learning algorithms based on a parallel rule-based system architecture. He advances the theme that inductive processes in such rule-based systems are a way of overcoming the *brittleness* of current AI systems, which is due to the narrow scope of their domain-specific knowledge.

In **chapter 21,** Berwick explores the issues of general constraints underlying processes of natural language acquisition. He discusses the relative importance of general, domain-independent learning principles versus domain-specific learning, and presents the *subset principle* for guiding generalization from positive-only examples.

Finally, in **chapter 22,** Silver describes a learning technique called *Precondition Analysis* that allows a program to learn strategies for problem solving. He illustrates his method with examples from the domain of algebraic equations.

The book concludes with a bibliography of research in machine learning since 1980, with a few major landmarks representing earlier research. (A comprehensive bibliography of previous research in this field can be found in *Machine Learning I.*) The bibliography is indexed by underlying learning strategy, domain of application, and research methodology. An updated glossary of terms in machine learning is also provided, as well as a bibliographical note about each author.

ACKNOWLEDGMENTS

This research was done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology, where the author worked while on leave of absence from the University of Illinois at Urbana-Champaign. The author wishes to thank Patrick Winston for inviting him to the Laboratory, for many fruitful and enjoyable discussions, and for comments on an earlier version of this chapter. The unique and stimulating environment of the MIT AI Laboratory and discussions at the Learning Group were helpful in shaping the ideas presented here. Support for the Laboratory's research is provided in part by the Advanced Research Projects Agency under the Office of Naval Research Contract N00014-80-C-0505.

The author thanks his coeditors, Jaime Carbonell and Tom Mitchell, for the collaboration and contributive comments. The discussion and remarks of Randy Davis from the MIT AI Laboratory were a valuable challenge and helped to improve this paper. He suggested the idea of using learning machines for protection against the misuse of learning machines. Richard Doyle, Michael Kashket, Boris Katz, and David Kirsh from the MIT AI Laboratory and Allan Collins from Bolt Beranek and Newman, Inc., provided important feedback and thoughtful comments on the earlier draft. The author expresses gratitude to Bob Stepp, Larry Rendell, Jeff Becker, Bruce Katz, and Brian Stout from the AI Laboratory at the University of Illinois for valuable suggestions and criticism. Useful comments and suggestions were provided by Ken Forbus from the UI Department of Computer Science. The author is indebted to Gail Thornburg from the UI Graduate School of Library and Information Science for her many insightful remarks and criticisms. Important suggestions were provided by

Jan Gorecki from the UI Department of Sociology. This work was supported in part by the National Science Foundation under Grant No. DCR-8406801 and the Office of Naval Research under Contract N00014-82-K-0186.

References

Andreae, P. M., "Constraint Limited Generalization: Acquiring Procedures from Examples," *Proceedings of AAAI-84*, Austin, Tex., pp. 6–10, 1984.

Berkson, W. and Wettersten, J., *Learning from Error*, Open Court, La Salle, Ill., 1984.

Buchanan, B. G., and Feigenbaum, E. A., "DENDRAL and Meta-DENDRAL: Their Applications Dimension," *Artificial Intelligence*, Vol. 11, pp. 5–24, 1978.

Buchanan, B. G., and Shortliffe, E. H. (Eds.), *Rule-based Expert Systems*, Addison-Wesley, Reading, Mass., 1984.

Caianiello, E. R., and Musso, G., *Cybernetic Systems: Recognition, Learning, Self-Organization*, Research Studies Press, Ltd., Letchworth, Hertfordshire, England; Wiley, New York, 1984.

Carbonell, J. G., Michalski, R. S., and Mitchell, T. M., "An Overview of Machine Learning," in *Machine Learning: An Artificial Intelligence Approach*, R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Eds.), Tioga, Palo Alto, Calif., 1983.

Cohen, P. R., and Feigenbaum, E. A. (Eds.), *The Handbook of Artificial Intelligence*, Vol. 3, Kaufmann, Los Altos, Calif., 1982.

Conrad, M., *Adaptability*, Plenum Press, New York, 1983.

Davis, R., and Lenat, D. B., *Knowledge-based Systems in Artificial Intelligence*, McGraw-Hill, New York, 1982.

Dietterich, T. G., and Michalski, R. S., "A Comparative Review of Selected Methods for Learning from Examples," in *Machine Learning: An Artificial Intelligence Approach*, R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Eds.), Tioga, Palo Alto, Calif., 1983.

Feigenbaum, A. E., Lecture at the First U.S.-China Joint Seminar on Automation and Intelligent Systems, Beijing, China, May 28–June 1, 1984.

Fogel, L., Owens, [illegible] and Walsh, M., *Artificial Intelligence Through Simulated Evolution*, Wiley, New York, [illegible]

Hayes-Roth, F., [illegible] Proofs and Refutations to Learn from Experience," in *Machine Learning: An Artificial In*[illegible] *Approach*, R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Eds.), Tioga, Palo Alto, [illegible] 1983.

Hayes-Roth, F., Waterman, D. A., and Lenat, D. B. (Eds.), *Building Expert Systems*, Addison-Wesley, Reading, Mass., 1983.

Hillis, W. D., "The Connection Machine (Computer Architecture for the New Wave)," AI Memo No. 646, MIT, September 1981.

Hinton, G. E., Sejnowski, T. J., and Ackley, D. H., "Boltzmann Machines: Constraint Satisfaction Networks That Learn," Technical Report CMU-CS-84-119, Department of Computer Science, Carnegie-Mellon University, 1984.

Hofstadter, D. R., *Godel, Escher, Bach: An Eternal Golden Braid*, Vintage, New York, 1980.

Holland, J., *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975.

Hume, D., *A Treatise of Human Nature*, L. S. Selby-Bigge (Ed.), Clarendon Press, Oxford, 1888.

Kawanobe, K., "Current Status and Future Plans of the Fifth Generation Computer Systems Project," *Proceedings of the International Conference on Fifth Generation Computer Systems*, COT, Tokyo, pp. 3–36, 1984.

Kuhn, T. S., *The Structure of Scientific Revolutions*, 2d ed. en., University of Chicago Press, Chicago, 1970.

Lakatos, I., "Falsification and the Methodology of Scientific Research Programmes," in *Criticism and the Growth of Knowledge*, A. Musgrave and I. Lakatos (Eds.), Cambridge University Press, Cambridge, 1970.

Langley, P., and Carbonell, J. G., "Approaches to Machine Learning," *Journal of the American Society for Information Science*, Vol. 35, No. 5, pp. 306–331, 1984.

Larkin, J., Reif, F., Carbonell, J., and Gugliotta, A., "FERMI: Flexible Expert Reasoning with Multi-Domain Inferencing," submitted to *Cognitive Science*, 1985.

Lenat, D. G., "The Role of Heuristics in Learning by Discovery: Three Case Studies," in *Machine Learning: An Artificial Intelligence Approach*, R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Eds.), Tioga, Palo Alto, Calif., 1983.

Marcus, M. P., *A Theory of Syntactic Recognition for Natural Language*, MIT Press, Cambridge, 1980.

McCarthy, J., "Programs with Common Sense," in *Semantic Information Processing*, M. Minsky (Ed.), MIT Press, Cambridge, 1968.

———, "President's Quarterly Message: AI Needs More Emphasis on Basic Research," *AI Magazine*, Vol. 4, pp. 4–5, 1983.

Medin, D. L., Wattenmaker, W. D., and Michalski, R. S., "Constraints in Inductive Learning: An Experimental Study Comparing Human and Machine Performance," submitted to *Cognitive Science*, 1985.

Michalski, R. S., "Variable-Valued Logic and Its Applications to Pattern Recognition and Machine Learning," in *Computer Science and Multiple-Valued Logic: Theory and Applications*, D. C. Rine (Ed.), North-Holland, 1975.

———, "Pattern Recognition as Rule-Guided Inductive Inference," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, No. 4, pp. 349–61, July 1980.

———, "A Theory and Methodology of Inductive Learning," *Artificial Intelligence*, Vol. 20, No. 2, pp. 111–161, 1983, 1983a.

———, "Learning Strategies and Automated Knowledge Acquisition: An Overview.:" in *Knowledge-Based Learning Systems*, L. Bolc, (Ed.), Springer-Verlag, 1985.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (Eds.), *Machine Learning: An Artificial Intelligence Approach*, Tioga, Palo Alto, Calif., 1983.

Michalski, R. S., Ko, H., and Chen, K., "Qualitative Process Prediction: A Method and a Program SPARC/G," *Reports of the Intelligent Systems Group*, ISG-12, Department of Computer Science, University of Illinois, Urbana, 1985.

Michie, D., *Machine Intelligence and Related Topics*, Gordon and Breach, New York, 1982.

Minsky, M. *The Society of Mind*, MIT Press, Cambridge (draft, April 1985), forthcoming.

Minsky, M., and Papert, S., *Perceptrons*, MIT Press, Cambridge, 1969.

Mitchell, T. M., Keller, R. M., and Kedar-Cabelli, S. T., "Explanation-Based Generalization: A Unifying View," *Machine Learning*, Vol. 1, No. 1 (Jan 1986): in press.

Mitchell, T. M., Utgoff, P. E., and Banerji, R., "Learning by Experimentation: Acquiring and Refining Problem-Solving Heuristics," in *Machine Learning: An Artificial Intelligence Approach*, R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Eds.), Tioga, Palo Alto, Calif., 1983.

Nilsson, N. J., *Learning Machines*, McGraw-Hill, New York, 1965.

Popper, K. R., *The Logic of Scientific Discovery*, Basic Books, New York, 1959.

———, *Objective Knowledge: An Evolutionary Approach*, rev. ed., Oxford University Press, Oxford, 1981.

Quinlan, J. R., "Discovering Rules from Large Collections of Examples: A Case Study," in *Expert Systems in the Microelectronics Age*, D. Michie (Ed.), Edinburg University Press, Edinburgh, 1979.

Rendell, L. A., "Toward a Unified Approach to Conceptual Knowledge Acquisition," *AI Magazine*, Vol. 4, pp. 19–27, Winter, 1983.

Robinson, J. A., "Logic Programming—Past, Present and Future," *New Generation Computing*, Vol. 1, No. 2, pp. 107–24, 1983.

Rosenblatt, F., "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychological Review*, Vol. 65, pp. 386–407, 1958.

Schank, R. C., *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*, Cambridge University Press, Cambridge, 1982.

———, "The Current State of AI: One Man's Opinion," *AI Magazine*, Vol. 4, No. 1, pp. 3–8, Winter Spring 1983.

Selfridge, M., "Pandemonium: A Paradigm for Learning," *Proceedings of the Symposium on Mechanization of Thought Processes*, D. Blake, and A. Uttley (Eds.), HMSO, London, pp. 511–29, 1959.

Simon, H. A., "Why Should Machines Learn?" in *Machine Learning: An Artificial Intelligence Approach*, R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Eds.), Tioga, Palo Alto, Calif., 1983.

Sleeman, D. H., "Inferring Student Models for Intelligent Computer-Aided Instruction," in *Machine Learning: An Artificial Intelligence Approach*, R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Eds.), Tioga, Palo Alto, Calif., 1983.

Sleeman, D. H., and Brown, J. S. (Eds.), *Intelligent Tutoring Systems*, Academic Press, New York, 1982.

Tsypkin, J. Z., *Foundations of the Theory of Learning Systems* (in Russian), Publisher Nauka, Moscow, 1972.

Weizenbaum, J., *Computer Power and Human Reason*, Freeman, San Francisco, 1976.

Winograd, T., "What Does It Mean to Understand Language?" in *Perspectives on Cognitive Science*, D. A. Norman (Ed.), Ablex, Norwood, N. J., 1981.

Winston, P. H., "Learning Structural Descriptions from Examples," in *The Psychology of Computer Vision*, P. H. Winston (Ed.), McGraw-Hill, New York, 1975.

______, "Learning and Reasoning by Analogy," *Communications of the ACM*, Vol. 19, No. 3, 1982.

______, *Artificial Intelligence*, 2d ed., Addison-Wesley, Reading, Mass., 1984.

Winston, P. H.; Binford, T. O.; Katz, B.; and Lowry, M. R., "Learning Physical Descriptions from Functional Definitions, Examples and Precedents," *Proceedings of the AAAI-83*, Washington, D.C., pp. 433–39, 1983.

Zagoruiko, N., "Empirical Prediction Algorithms," in *Computer Oriented Learning Processes*, J. C. Simon (Ed.), Noordhoff, Leyden, 1976.