Goal-Driven Learning*

Ashwin Ram
College of Computing
Georgia Institute of Technology

David B. Leake
Computer Science Department
Indiana University

*MIT Press/Bradford Books, Cambridge, MA, in press.

Table of Contents

Copyright

The following chapters are reprinted with permission:

Chapter 2, L.E. Hunter, "Planning to Learn," is reprinted from *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, pages 26–34, 1990. ©1990 Cognitive Science Society, Inc., used with permission.

Chapter 3, S. Minton, "Quantitative Results Concerning the Utility of Explanation-Based Learning," is reprinted from *Artificial Intelligence*, 42:363–391, 1990. ©1990 by Elsevier Science Publishers, B.V.

Chapter 4, A. Ram and L. Hunter, "The Use of Explicit Goals for Knowledge to Guide Inference and Learning," is reprinted from *Applied Intelligence*, 2(1):47–73, 1992. ©1992 by Kluwer Academic Publishers.

Chapter 5, L.W. Barsalou, "Deriving Categories to Achieve Goals," is reprinted from *The Psychology of Learning and Motivation: Advances in Research and Theory*, Volume 27, G.H. Bower, editor, 1991. ©1991 by Academic Press, Inc.

Chapter 6, E.J. Wisniewski and D.L. Medin, "Harpoons and Long Sticks: The Interaction of Theory and Similarity in Rule Induction," is reprinted from D. Fisher and M.J. Pazzani (eds.), *Concept Formation: Knowledge and Experience in Unsupervised Learning*, 1991. ©1991 by Morgan Kaufmann Publishers, Inc.

Chapter 7, A. Ram and M.T. Cox, "Introspective Reasoning using Meta-Explanations for Multistrategy Learning," is reprinted with minor revisions from *Machine Learning IV: A Multistrategy Approach*, R.S. Michalski and G. Tecuci, editors, 1994. ©1994 by Morgan Kaufmann Publishers, Inc.

Chapter 8, M. desJardins, "Goal-Directed Learning: A Decision-Theoretic Model for Deciding What to Learn Next," is reprinted with minor revisions from *Proceedings of the ML-92 Workshop on Machine Discovery*, Ninth International Machine Learning Conference, pages 147–151, 1992. ©1992 by M. desJardins.

Chapter 9, D. Leake, "Goal-Based Explanation Evaluation," is reprinted from *Cognitive Science*, 15(4):509–545, 1991. ©1991 by Ablex Publishing Corporation. Reprinted with permission from Ablex Publishing Corporation.

Chapter 10, L. Pryor and G. Collins, "Planning to Perceive," is reprinted from *Proceedings of the 1992 AAAI Spring Symposium on Selective Perception*, Stanford, CA. ©1992 by L. Pryor and G. Collins.

Chapter 12, X. Xia and D.-Y. Yeung, "A Learning Model for the Selection of Problem Solving Strategies in Continuous Physical Systems," is reprinted from

Proceedings of the Fourth IEEE Conference on Artificial Intelligence Applications, pages 200–206, San Diego, CA, 1988. ©1988 by IEEE.

Chapter 13, D. Gordon and D. Perlis, "Explicitly Biased Generalization," is reprinted with minor revisions from *Computational Intelligence*, 5(2):67–81, 1989. ©1989 by National Research Council of Canada.

Chapter 14, E. Ng and C. Bereiter, "Three Levels of Goal Orientation in Learning," is reprinted from *The Journal of the Learning Sciences*, 1(3&4):243–271, 1991. ©1991 by Lawrence Erlbaum Associates, Inc.

Chapter 15, J.J.A. van Berkum, H. Hijne, T. de Jong, W.R. van Joolingen, and M. Njoo, "Characterising the Application of Computer Simulations in Education: Instructional Criteria," is reprinted with minor revisions from "Aspects of Computer Simulations in an Instructional Context," *Education and Computing*, 6:231–239, 1991. ©1991 by SAFE.

Chapter 16, D.B. Leake and A. Ram, "Goal-Driven Learning: Fundamental Issues and Symposium Report," is reprinted with minor revisions from an article in *AI Magazine*, 14(4). ©1993 by the American Association for Artificial Intelligence.

Preface by Tom Mitchell

Understanding the process of learning, and giving this capability to computers, is among the most ambitious goals of modern computer science. The past decade has seen research progress toward this goal along several fronts: developing new learning algorithms, modelling human learning processes, and theoretically characterizing the complexity of learning.

Among the most significant new themes over this decade is the exploration of how the learner's goals and prior knowledge drive the learning process. Before 1980, the vast majority of research on learning focussed on the task of estimating or hypothesizing an unknown function given only a sample of its inputs and outputs, with no explicit notion of the learning goal. This research grew out of the tradition of earlier work on statistical pattern recognition, and early studies in psychology on learning "nonsense" concepts such as abstract geometric patterns. But over time many researchers became convinced that this was too simple a formulation of the learning problem to model the surprising learning skills exhibited by humans.

To see the issue here, consider learning to play a game such as chess. Suppose you have just lost your queen, and wish to learn the general pattern of board pieces that led to this failure. Inductive learning methods require hundreds or thousands of such failures (and non-failures), to hypothesize which of the many board features distinguish the positions in which you lose your queen from those in which you do not. But people are typically able to learn such concepts from only a handful of examples. How? People appear to direct learning towards concepts relevant to their goals – by explaining the cause of their failure, and thereby noticing the relevant board features (e.g., the opposing knight that was simultaneously threatening your King and Queen) while ignoring the irrelevant (e.g., the three pawns in the second row). The oversimplification of early work on learning was that it omitted any reference to the learner's goals, and therefore could not model this kind of learning by explaining.

During the 1980's researchers began exploring the role of explanations, goals, and explicit prior knowledge in the learning process. Explanation-based learning algorithms were developed that generalized more accurately than earlier inductive approaches, by explicitly taking into account the learning goal (e.g., to avoid losing the queen) and related prior knowledge (e.g., the legal moves of chess). The key insight underlying this work was that learning is much easier to understand and duplicate in computers if the learning goal and related knowledge are explicitly manipulated by the learning algorithm.

This initial work led to a flurry of research on goal-driven learning, exploring more broadly the ways in which learning processes are influenced by learning goals. Whereas initial research considered how learning goals influence the process of generalizing from examples, more recent work has considered how learning goals drive other processes as well, such as the process of experimenting to collect new training data, and the process of generating useful learning subgoals. Whereas the initial research produced simplified, brittle algorithms for using the learning goal to guide generalization, more recent work has produced significantly more robust and practical methods and has raised questions concerning the origins of learning goals, the role of goals in guiding other learning tasks, and how to choose appropirate learning strategies to achieve learning goals.

Current research in goal-driven learning deals with a wide range of issues dealing with how and when learning goals arise and the ways in which goals influence a broad range of learning processes. These issues, which are now being addressed in machine learning, cognitive psychology, and education, are the focus of this volume, which summarizes recent work on goal-driven learning, and presents a number of new research results in this area. As you will see from many of the chapters, the field is actively exploring new directions, and many of the approaches are still under development. Although we still lack a full understanding of how to best use goals to guide the learning process, it seems obvious that if we are to progress in understanding learning, then we must take into account more and more of the the rich context in which it occurs. Taking learning goals into account is one essential step, and it is difficult to imagine a future for machine learning or cognitive science in which this does not play a central role.

Professor Tom M. Mitchell Carnegie Mellon University Pittsburgh, PA

Editors' Preface

In cognitive science, artificial intelligence, psychology, and education, a growing body of research supports the view that the learning process is strongly influenced by the learner's goals. Several experimental studies have shown that people with different goals process information differently. Studies in educational contexts have shown that different types of goals influence learning in different ways, and have attempted to use this insight in the design of effective educational environments. The importance of learner goals is supported by computational machine learning models, which provide functional arguments for goal-based focusing of learner effort, and from psychological evidence for the importance of student goals in educational settings. Investigators in each of these areas have independently pursued the common issues of how learning goals arise, how they affect learner decisions of when and what to learn, and how they guide the learning process.

The fundamental tenet of goal-driven learning is that learning is largely an active and strategic process in which the learner, human or machine, attempts to identify and satisfy its information needs in the context of its tasks and goals, its prior knowledge, its capabilities, and environmental opportunities for learning. It is increasingly evident that investigation of goal-driven learning can benefit from a multidisciplinary effort employing diverse perspectives on a common research agenda. To this point, however, research in goal-driven learning has largely been confined to isolated efforts, with little framework to connect related results and to aid in their analysis. The purpose of this book is to establish such a framework, to collect and solidify existing results on goal-driven learning, and to point the way for future investigations of goal-driven learning.

The book begins with a discussion of fundamental questions for goal-driven learning: the motivations for adopting a goal-driven model of learning, the basic goal-driven learning framework, the specific issues raised by the framework that a theory of goal-driven learning must address, the types of goals that can influence learning, the types of influences those goals can have on learning, and the pragmatic implications of the goal-driven learning model (chapter 1). The remainder of the book is divided into two parts. The first is a collection of recent research papers that serve as case studies in goal-driven learning. Each paper addresses a piece of the goal-driven learning puzzle, reflecting a particular research perspective from one of the several disciplines that have been investigating this area in recent years. These works address issues such as the justification of goal-driven learning models through functional arguments about the role and utility of goals in learning (chapters 2–4), the justification of such models through cognitive results (chapters 5, 6, 14), goal-based processes for deciding what to learn (chapters 7, 8) and for guiding learning and the learning process (chapters 4, 7, 9–13), and pragmatic implications of goal-driven learning for design of instructional environments (chapters 14, 15).

The second part of the book is based on the Symposium on Goal-Driven Learning organized by David Leake and Ashwin Ram at the Fourteenth Annual Conference of the Cognitive Science Society in Bloomington, Indiana, in 1992. It presents an overview of the workshop discussion and a collection of papers from the symposium panelists representing their individual perspectives on fundamental issues and their proposals for fruitful future directions in goal-driven learning research.

The works in this volume reflect both the diversity of goal-driven learning research and the fundamental relationship of different approaches within the broader goal-driven learning framework. Together, they provide a comprehensive overview of recent research in goal-driven learning and illuminate on-going investigations and open issues to provide a foundation for future study of goal-driven learning.

About the editors

Ashwin Ram received his B.Tech. in Electrical Engineering from the Indian Institute of Technology, New Delhi, in 1982, and his M.S. in Computer Science from the University of Illinois at Urbana-Champaign in 1984. He received his Ph.D. degree from Yale University for his dissertation on "Question-Driven Understanding: An Integrated Theory of Story Understanding, Memory and Learning" in 1989. He is currently an Assistant Professor in the College of Computing of the Georgia Institute of Technology, and an Adjunct Professor in the School of Psychology. His research interests lie in the areas of machine learning, natural language understanding, explanation, and cognitive science, and he has many research publications in these areas. His address is: College of Computing, Georgia Institute of Technology, Atlanta, Georgia 30332-0280. (ashwin.ram@cc.gatech.edu)

David Leake received his B.A. with high honors in Mathematics from Haverford College in 1980, and his A.M. in Mathematics from Brown University in 1984. In 1990 he received his Ph.D. in Computer Science from Yale University. He is currently an Assistant Professor of Computer Science and a member of the Cognitive Science faculty at Indiana University. His primary interests are case-based reasoning, explanation, and memory. He is the author of the book *Evaluating Explanations: A Content Theory* (Lawrence Erlbaum Associates, 1992). His address is: Computer Science Department, Indiana University, Bloomington, IN 47405. (leake@cs.indiana.edu)

The editors would like to take this opportunity to note that the order of their names on this work was arbitrarily determined by the flip of a coin.

List of Contributors

Lawrence Barsalou received his Ph.D. in Psychology from Stanford University in 1981. He has held faculty positions at Emory University, the Georgia Institute of Technology, and is currently a Professor of Psychology at the University of Chicago. His research interests include issues in knowledge, thought, memory, language, and computational models of cognition. His address is: Department of Psychology, 5848 South University Avenue, The University of Chicago, Chicago, IL 60637. (l-barsalou@uchicago.edu).

Carl Bereiter is a professor in the Centre for Applied Cognitive Science at the Ontario Institute for Studies in Education. With Marlene Scardamalia, he has published *The Psychology of Written Communication* (Erlbaum, 1987) based on their research on mature and immature writing strategies, and more recently, "Surpassing Ourselves: An Inquiry Into the Nature and Implications of Expertise." His address is: The Ontario Institute for Studies in Education, 252 Bloor Street West, Toronto, Ontario, M5S 1V6 Canada. (c_bereiter@utoroise.bitnet).

Jos van Berkum received his degree in Experimental and Cognitive Psychology from the University of Groningen in 1988. He is currently investigating the representation and processing of syntactic gender at the Max Planck Institute for Psycholinguistics, P.O. Box 310, NL-6500 AH Nijmegen, The Netherlands. (berkum@mpi.nl).

Jaime G. Carbonell is a Professor of Computer Science and the Director of the Center for Machine Translation at Carnegie-Mellon University. He received his Ph.D. degree in Computer Science from Yale University in 1976. He has edited or authored several books, four of them in machine learning and one in machine translation. He has served as chairman of SIGART (1983-1985), the special interest group on AI of the ACM, and is a founder and director of Carnegie Group Inc. His research interests span several areas of artificial intelligence, including: knowledge-based machine translation, machine learning, natural language processing, planning and problem solving. His address is: School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213. (jgc@cs.cmu.edu).

Gregg Collins received his Ph.D. from Yale University in 1987. His research interests are in planning and machine learning, with particular emphasis on explanation-based learning, case-based reasoning, and model-based reasoning. He is currently an Assistant Professor of Electrical Engineering and Computer Science at the Institute for the Learning Sciences, Northwestern University, 1890 Maple Avenue, Evanston, IL 60201. (collins@ils.nwu.edu).

Michael Cox received a B.S. in Information and Computer Science with highest honors in 1986 from the Georgia Institute of Technology, where he is currently a Ph.D. candidate. His research interests include human self-understanding and its relation to learning and awareness, memory, and forgetting. His address is: College of Computing, Georgia Institute of Technology, Atlanta, Georgia 30332. (cox@cc.gatech.edu).

Marie desJardins received her Ph.D. in Computer Science from the University of California at Berkeley in 1992. Her research interests include machine learning, planning, reasoning and decision making under uncertainty, and intelligent tutoring systems. Her address is: SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025. (marie@erg.sri.com).

Oren Etzioni received his Ph.D. from Carnegie Mellon University in 1991. His research interests include software agents, machine learning, planning, plan recognition, and human-computer interaction. In 1991 he launched the Internet Softbots project. In 1993, he received an NSF Young Investigator Award. He is currently an Assistant Professor in the Department of Computer Science and Engineering at the University of Washington, Seattle, WA 98195. (etzioni@cs.washington.edu).

Yolanda Gil received her Ph.D. in Computer Science from Carnegie Mellon University in 1992. She is currently a research scientist at the Information Sciences Institute of the University of Southern California, and a research assistant professor in the Computer Science Department. Her current interests are planning, learning, and knowledge acquisition. Her address is: University of Southern California, Information Sciences Institute, 4676 Admiralty Way, Marina del Rey, CA 90292. (gil@doc.isi.edu).

Diana Gordon received her Ph.D. in Computer Science from the University of Maryland in 1990. Her research interests include the evaluation of biases in machine learning, active learning, problem reformulation, and agent knowledge acquisition. She is currently a research scientist at the Naval Research Laboratory, Code 5514, 4555 Overlook Avenue S.W., Washington, DC 20375-5337. (gordon@aic.nrl.navy.mil).

Hans Hijne studied Cognitive Psychology at the University of Amsterdam. He has participated in several projects on intelligent help systems and simulation environments for training and education. He is now a free lance consultant for computer-based training applications.

Lawrence Hunter is the Director of the Machine Learning Project at the National Library of Medicine. He received his Ph.D. in Computer Science in 1989 from Yale University. His research involves the goal-driven selection, combination and application of multiple machine learning technologies to problems in medicine and molecular biology, the origins and implications of desires for knowledge, the evolution of cognitive abilities, and the social implications of technical advances in biology and computer science. He can be reached at the National Library of Medicine, Lister Hill Center, Building 38A, Mail Stop 54, Bethesda, MD 20894. (hunter@nlm.nih.gov).

Ton de Jong received his Ph.D. in Cognitive Psychology from the Eindhoven University of Technology. He has been affiliated with the Delft University of Technology, the University of Amsterdam, and the Eindhoven University of Technology, and is currently an Associate Professor at the University of Twente. His address is: Faculty of Educational Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. (jong@edte.utwente.nl).

Wouter van Joolingen received his Ph.D. from Eindhoven University of Technology in 1993. He has participated in several projects on simulation learning environments for education, and is now working part-time for the Faculty of Educational Research and Technology at the University of Twente. He can be reached at the Eindhoven University of Technology, Department of Philosophy and Social Science, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. (w.r.v.joolingen@wmw.tue.nl).

Robert Joseph received his Ph.D. from Carnegie Mellon University and is currently a Member of the Technical Staff at U.S. West Technologies, where he is working on broadband applications (i.e., video-on-demand, home shopping). He is the Third Vice President for the National Association of Black Telecommunications Professionals and is on the Editorial Advisory Board of the International Journal of Engineering Intelligent Systems. His research interests include intelligent agents and data visualization. His address is U.S. West Technologies, 4001 Discovery Drive, Boulder, CO 80303. (rjoseph@advtech.uswest.com).

Craig Knoblock is a Senior Research Scientist at the USC Information Sciences Institute and a Research Assistant Professor in the Computer Science Department at the University of Southern California. He received his Ph.D. in Computer Science from Carnegie Mellon University in 1991. His current research interests are on developing and applying planning, machine learning and knowledge representation techniques to the problem of providing access to distributed heterogeneous information sources. His address is: University of Southern California, Information Sciences Institute, 4676 Admiralty Way, Marina del Rey, CA 90292. (knoblock@isi.edu).

Douglas Medin received his Ph.D. in Psychology from the University of South Dakota in 1968. He has been a faculty member at Rockefeller University, the University of Illinois, and the University of Michigan, and is currently a Professor of Psychology at Northwestern University, Psychology Department, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208. (medin@nwu.edu).

Ryszard Michalski Ryszard S. Michalski is a Chaired Professor of Computer Science and Systems Engineering, and Director of the Center for Artificial Intelligence, at George Mason University. Prior to joining George Mason University he was Professsor of Computer Science at the University of Illinois at Champaign-Urbana. He co-edited volumes I-IV of the book series *Machine Learning;* his current research areas include machine learning and inference, knowledge discovery, integrated learning and vision systems, cognitive modelling, and practical applications of intelligent systems. His current address is: Center for Artificial Intelligence, George Mason University, Fairfax, VA 22030-4444 (michalski@aic.gmu.edu).

Elijah Millgram received his Ph.D. from Harvard University. His current research studies the ways in which agents reason about goal selection and on the role of arguments from computational complexity in theories of rationality. He is currently teaching in the Philosophy Department, Princeton University, Princeton, NJ 08544. (lije@clarity.princeton.edu).

Steven Minton received his Ph.D. in Computer Science from Carnegie Mellon University in 1988. He is presently working in the areas of machine learning, planning, and scheduling in the Artificial Intelligence Research Branch at the NASA Ames Research Center, Recom Technologies, Mail Stop 269-2, Moffett Field, CA 94035. (minton@ptolemy.arc.nasa.gov).

S. Narayanan received M.S. degrees in Industrial Engineering from the University of Alabama in 1989 and from the Georgia Institute of Technology in 1991. His research interests involve computational modeling of complex systems, object-oriented simulation, and the application of artificial intelligence techniques to design decision aids in integrated manufacturing systems. He is currently a Ph.D. candidate at the Center for Human-Machines Systems Research, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332. (sn@chmsr.isye.gatech.edu).

Evelyn Ng received a Ph.D. in Education from the University of Toronto. She has served on the faculty at Simon Fraser University, and is now a consultant with Ng and Prosser Associates, a research and program evaluation firm. She is interested in applied cognitive psychology and computer-assisted language learning. Her current address is: 6636 Sumas Drive, Burnaby, British Columbia, V5B 2V2 Canada.

Melanie Njoo received her degree in Educational Psychology at the University of Leiden. She worked at the Eindhoven University of Technology on exploratory learning with simulation-based learning environments, and now designs and develops instructional systems and support for the implementation of new information technology at BSO/Instruction Technology, P.O. Box 543, 3740 AM Baarn, The Netherlands.

Donald Perlis is an Associate Professor of Computer Science at the University of Maryland in College Park, and also has an appointment in the Institute for Advanced Computer Studies. His research interests include foundations of commonsense reasoning, cognitive modelling, and space-and-time situated reasoning and acting. He has a Ph.D. in Computer Science from the University of Rochester and a Ph.D. in Mathematics from New York University. He can be reached at the Computer Science Department, University of Maryland, College Park, MD 20742. (perlis@cs.umd.edu).

Louise Pryor received her Ph.D. from Northwestern University in 1993. Her research interests lie in the areas of planning and plan execution, with particular emphasis on the problems facing an agent operating in an unpredictable world. She is a Lecturer in Computer Science at the University of Birmingham, School of Computer Science, Edgbaston, Birmingham B15 2TT, England. (l.m.pryor@cs.bham.ac.uk).

Paul Thagard is a Professor of Philosophy and Adjunct Professor of Psychology and Computer Science at the University of Waterloo. He received the Ph.D. in Philosophy from the University of Toronto in 1977, and his M.S. in Computer Science from the University of Michigan in 1985. He can be reached at the Philosophy Department, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. (prthagar@logos.uwaterloo.ca).

Manuela Veloso is an Assistant Professor of Computer Science at Carnegie Mellon University. She received her Ph.D. degree in Computer Science from Carnegie Mellon University in 1992. Her Ph.D. dissertation research was on learning by analogical reasoning in general problem solving. Her current research interests include planning and learning and the integration of machine vision and machine learning. Her address is School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213. (mmv@cs.cmu.edu).

Edward Wisniewski received his Ph.D. in Cognitive Science in 1989 from Brown University. His research interests include cognitive science, human and machine learning, and conceptual combination. He is currently an Assistant Professor in the Psychology Department, Northwestern University, Evanston, IL 60208. (ed_wisniewski@psych.nwu.edu).

Don (Xiaodong) Xia conducted his graduate work in intelligent systems at University of Southern California, and is currently a member of the senior research staff at Telecom Australia Research Laboratories, where he is working on advanced software techniques, including AI and distributed systems, with an emphasis on their applications to telecommunications software systems. He can be reached at Telecom Australia Research Labs, 770 Blackburn Road – M6, Clayton, Victoria 3168, Australia. (d.xia@trl.oz.au).

Dit-Yan Yeung received his Ph.D. in Computer Science in 1989 from the University of Southern California. His research interests include artificial neural networks and pattern recognition. He has served on the faculty of the Illinois Institute of Technology in Chicago and is now at the Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. (dyyeung@cs.ust.hk).

Chapter 1

Learning, Goals, and Learning Goals

Ashwin Ram and David B. Leake

1 Why goals?

In cognitive science, artificial intelligence, psychology, and education, a growing body of research supports the view that learning is largely a goal-directed process. Experimental studies show that people with different goals process information differently; work in machine learning presents functional arguments for goal-based focusing of learner effort. Recent work in these fields has focussed on issues of how learning goals arise, how they affect learner decisions of when and what to learn, and how they guide the learning process. It is increasingly evident that investigation of goal-driven learning can benefit from bringing these perspectives together in a multidisciplinary effort (Leake & Ram, 1993).

The central idea underlying goal-driven learning is that, because the value of learning depends on how well the learning contributes to achieving the learner's goals, the learning process should be guided by reasoning about the information that is needed to serve those goals. The effectiveness of goal-driven learning depends on being able to make good decisions about when and what to learn, on selecting appropriate strategies for achieving the desired learning, and on guiding the application of the chosen strategies. Research into such topics includes the development of computational models for goal-driven learning, the testing of those models through psychological experiments and empirical experiments with computer programs, the justification of the models through functional arguments about the role and utility of goals in learning, and the use of models of goal-driven learning in guiding the design of educational environments. The common themes in these research efforts are the investigation of types of learning goals, the origins of learning goals, and the role of goals in the learning process.

Research on goal-driven learning in artificial intelligence has been motivated largely by computational arguments. The problem of combinatorial explosion of inferences is well known; in any realistic task domain, time and resource constraints prohibit consideration of all but a few of the possible inferential paths. Consequently, any reasoner, human or machine, must focus its attention and resources on pursuing those inferential paths that are likely to be most useful. Similarly, in any realistic situation, there are several different types of learning that a reasoner might perform, several kinds of new knowledge that a reasoner might acquire, and several kinds of reformulation or reorganization of existing knowledge that a reasoner might carry out. Again, due to time and resource constraints it is only practical to perform a few of these operations. Consequently, the reasoner must focus its attention and resources on executing the learning operations that are likely to be most useful. Because the utility of an inference or a piece of knowledge can best be evaluated relative to a particular task or goal, goal-based considerations must guide reasoning and learning.

In addition to these computational arguments for goal-driven learning, research in goal-driven learning has a cognitive basis in psychological research. This research has established much evidence for the influence of goals and beliefs on human learning, and for the use of active, strategic, and goal-driven processes in many kinds of learning that humans perform. However, many questions remain concerning the kinds of goals that people pursue, the conditions under which those goals influence learning, and the kinds of learning that are influenced by those goals.

Research in cognitive science combines the cognitive perspective of psychology with the computational perspective of artificial intelligence, developing computational models of human learning that are evaluated using computational metrics as well as by comparison with human performance. Research in education has also been concerned with psychological data about human learning, but from a pragmatic perspective. This research has attempted to use empirical evidence to guide the design of instructional and educational scenarios so as to facilitate learning, taking as its starting point the evidence for facilitation of certain kinds of learning by particular kinds of goals. These scenarios have also been used as the basis for further psychological experimentation to validate the underlying theories. In this chapter, we describe a framework for goal-driven learning and its relationship to prior and current theories from each of these perspectives.

2 An everyday example

Goal-driven learning is triggered when a reasoner needs to learn in order to improve its performance at some task. A goal-driven learner determines *what* to learn by reasoning about the information it needs, and determines *how* to learn by reasoning about the relative merit of alternative learning strategies in the current circumstances. For example, for a first-time stereo buyer, the goal of getting good buy on a stereo may give rise to at least two learning goals: a goal to learn the best sources for sound equipment and to a goal to learn how to judge the merits of competing equipment. Each of these learning goals may trigger learning subgoals. In order to learn the best place to buy sound equipment, the buyer may first have to learn general criteria for what constitutes a good store for buying sound equipment, and then specifics about prices, service, etc. to classify different stores. In order to learn how to judge particular equipment, the buyer will have to learn about the classes of alternatives available and about specific equipment within those classes. Thus some learning goals involve gathering information in the external world, while others involve reformulating or changing information that is already known, by operations such as forming generalizations or reorganizing memory.

In order to perform the desired learning, the stereo buyer must select strategies for accomplishing each of its learning goals. For example, the buyer may choose between learning strategies including asking others' opinions, reading magazine articles, forming inductive or explanation-based generalizations from demonstrations of equipment, or even disassembling equipment to determine the quality of its electronic components. Learning strategy selection depends on factors such as the buyer's prior knowledge, the buyer's resources (e.g., how much time the buyer can spend on the shopping process), opportunities (e.g., happening to meet an expert on sound equipment at a party), and the buyer's own abilities (e.g., whether the buyer has the expertise to judge the quality of equipment by disassembling it).

This example illustrates the value of goal-driven learning in focusing learner effort, and also suggests the range of roles that goals can play in influencing learning. Goals determine how much effort to allocate to performance tasks (e.g., the task of buying a stereo), indirectly influencing the resources available for the learning that will be performed as part of that task. Goals also determine the focus of attention when new information is received as input (e.g., focusing attention on announcements of stereo sales). They determine what should be learned (e.g., determining that it is worthwhile to generalize about relationships between store types and prices). They give criteria for evaluating the results of learning and deciding what learned information to store (in this example, the value of learning is its usefulness for guiding the shopping decision). Table 1 summarizes these and other possible roles of goals in learning. In the following sections we concentrate on developing a framework and terminology on which to base our analysis of goal-driven learning, and after developing that framework we return to the ways that goals affect learning in section 8.

3 Towards a planful model of learning

As the previous example illustrates, a goal-driven learner makes decisions about what, how, and when to learn in order to further its goals. In this view, learning can be considered a "planful" process (e.g., Etzioni, Hanks, Weld, Draper, Lesh, & Williamson, 1992; Hunter, 1990/chapter 2; Leake/chapter 20; Michalski &

Guiding the performance task by:

- Determining the resources made available to the performance task
- Guiding the control or search procedure used in the performance task
- Guiding retrieval of plans, problem solutions, and other types of knowledge
- Focusing attention on certain aspects of the input
- Guiding the evaluation of the outcome of the performance task

Guiding the learning task, by:

- Specifying the target of learning (desired output of a learning algorithm)
- Selecting the learning algorithms to be used
- Constraining the learning process (for example, influencing the policies under which the learning algorithms operate)
- Focussing the search for information needed to carry out the learning
- Determining when learning should be attempted
- Aiding evaluation of results of learning with respect to the desired output

Guiding storage, by:

- Selecting what to store
- Determining how learned knowledge is indexed

Table 1: Ways in which goals can influence learning.

Ram, chapter 21; Pryor & Collins, 1992/chapter 10; Ram & Cox, 1994/chapter 7; Ram, Cox, & Narayanan, chapter 18; Ram & Hunter, 1992/chapter 4; Redmond, 1992; Quilici, in press; Schank & Abelson, 1977; Xia and Yeung, 1988/chapter 12). This learning process is analogous to models of problem solving in which the reasoner uses task goals to formulate action plans for achieving these goals (e.g., Newell & Simon, 1972; Greeno & Simon, 1988; VanLehn, 1989). Learning actions or schemas are selected, combined, and invoked appropriately on the basis of existing learning goals and available environmental opportunities for learning. Learning is a behavior explicitly carried out to seek information, driven by needs arising from the reasoner's performance on a task that learning is intended to facilitate, and mediated by the formulation and manipulation of explicit learning goals.

The motivation for the goal-driven approach is to control processing in a rich world. Simply put, knowledge that is valid in principle need not necessarily be useful (Mitchell & Keller, 1983); thus, it is desirable to avoid the effort involved in learning knowledge that does not contribute to the reasoner's overall purpose. For example, Ram and Hunter (1992/chapter 4) argue that, due to the computational complexity of reasoning about the combinatorially large number of inferences that are possible in any realistic situation, it is essential to focus inferential and learner effort on deriving those pieces of knowledge that are likely to be most useful. Hunter (1990/chapter 2) argues that inference during learning (such as inductive inference) is also potentially combinatorially explosive and that explicit consideration of desirable knowledge should be used to guide this inference. Likewise, Leake (1992) argues for similar reasons that decisions about what to learn about new situations must be driven by characterizations of the learner's information needs. Theoretical analyses (desJardins, 1992/chapter 8; Etzioni, 1992; Francis & Ram, 1993; Gratch & DeJong, 1993), as well as empirical investigations of the utility of learning (Minton, 1990/chapter 3; Tambe, Newell, & Rosenbloom, 1990) provide support for this argument. Active, goal-driven learning implies the ability to make explicit decisions about what, when, and how to learn (Ram, Cox, & Narayanan/chapter 18). Thus some of the motivations for goal-based approaches include (see also Cox & Ram, 1994):

- **Alleviating problems of computational complexity:** The ability of a reasoner to make decisions about its reasoning and learning processes helps to alleviate problems caused by the computational complexity of reasoning in an open world, by enabling the reasoner to focus its efforts towards processing that serves its goals (Cox, 1993; Hunter, 1990/chapter 2; Leake, 1992; Leake/chapter 20; Ram & Hunter, 1992/chapter 4). An analysis of the utility of learning can help in determining the target of learning (desJardins, 1992/chapter 8), in guiding learning processes (Gratch & DeJong, 1993; Gratch, DeJong, & Chien, 1994; Provost, 1994), and also in deciding whether to learn at all (Markovitch & Scott, 1993; Minton, 1990/chapter 3).

- **Facilitating the use of opportunities to learn:** If a reasoner does not have sufficient resources at the time it realizes it has a need to learn, or if the requisite knowledge is not available at that time, the reasoner can suspend its learning goals in memory so that they can be retrieved and pursued at a later time (Hunter, 1990/chapter 2; Hammond, Converse, Marks, & Seifert, 1993; Ram, 1991, 1993; Ram, Cox, & Narayanan/chapter 18; Ram & Hunter, 1992/chapter 4).

- **Improving the global effectiveness of learning:** Taking goal priorities and goal dependencies into account when deciding what to learn and how to coordinate multiple learning strategies improves the effectiveness of learning in a system with multiple goals. Learning strategies, represented as methods for achieving learning goals, can be chained, composed, and optimized, resulting in learning plans that are created dynamically and pursued in a flexible manner (Cox, 1993; Cox & Ram, 1994; Gratch, DeJong, & Chien, 1994; Hadzikadic & Yun, 1988; Hunter, 1990/chapter 2; Michalski, 1993; Michalski & Ram, chapter 21; Ram & Hunter, 1992/chapter 4; Redmond, 1992; Stroulia & Goel, 1994).

- **Increasing the flexibility of learning:** In situations involving multiple reasoning failures, multiple active and suspended learning goals, multiple applicable learning strategies, and limited resources, direct mapping from specific types of failures to individual learning strategies is impossible, and an active, planful approach becomes necessary. For a given failure, there may be more than one algorithm which needs to be applied for successful learning and, conversely, a given algorithm may apply to many different types of failures (Cox, 1993; Cox & Ram, 1994; Krulwich, Birnbaum, & Collins, 1993; Ram, Cox, & Narayanan/chapter 18). A planful model of learning allows decoupling of many-to-many relationships, leading to more flexible behavior (Cox, 1993, Cox & Ram, 1994).

- **Improving management of interactions between learning processes:** Explicit formulation of learning goals facilitates detection of dependency relationships, so that goal violations can be avoided (Cox, 1993, Cox & Ram, 1994). When multiple items are learned from a single episode, the changes resulting from one learning algorithm may affect the knowledge structures used by another algorithm. Such dependencies destroy any implicit assumption of independence built into a given learning algorithm that is used in isolation. For example, one learning algorithm may split a concept definition into separate schemas or otherwise modify the definition. Therefore, an indexing algorithm that uses the attributes of concepts to create indices must necessarily follow the execution of any algorithm that changes the conceptual definition.

Psychological evidence also supports the existence of goal-based influences on human focus of attention, inference, and learning (e.g., Barsalou, 1991/chapter 5; Faries & Reiser, 1988 Hoffman, Mischel, & Mazze, 1981; Ng & Bereiter, 1991/chapter 14; Seifert 1988, Srull & Wyer, 1986; Wisniewski & Medin, 1991/chapter 6; Zukier, 1986; see also discussion by Hunter, 1990/chapter 2). These ideas are related to the "goal satisfaction principle" of Hayes-Roth and Lesser (1976), which states that more processing should be given to knowledge sources whose responses are most likely to satisfy processing goals, and to the "relevance principle" of Sperber and Wilson (1986), which states that humans pay attention only to information that seems relevant to them. Those principles make sense because cognitive processes are geared to achieving a large cognitive effect for a small effort. To achieve this, the understander must focus its attention on what seems to it to be the most relevant information available. Goals can facilitate learning even when they are not generated internally by the reasoner; for example, Steinbart (1992) shows that asking users questions (i.e., "creating" knowledge goals in people) can help them learn from a computer-assisted training program, and Patalano, Seifert, and Hammond (1993) show that presenting users with a goal and a plan to achieve it can facilitate later detection of relevant features of a situation. There is also much research on the origins of goals; for example, Graesser, Person, and Huber (1992) discuss several types of questions, or goals to seek information, and the cognitive mechanisms that generate them. Many of these are related to the learning goal formulation mechanisms discussed here.

The goal-driven learning framework does not imply that all processing is explicitly goal-driven. A reasoner that was completely goal-driven would only notice what it was looking for already; it would not be able to respond to and learn from unexpected input. Instead, it is reasonable to assume that there would be some automatic, bottom-up, or non-goal-driven processing during reasoning and learning, which would support strategic, top-down, or goal-driven processes such as those discussed here (e.g., Barsalou/chapter 17; Kintsch, 1988; Leake, 1992; McKoon & Ratcliff, 1992; Ram, 1991).

A significant body of psychological research points to the influence of "metacognition"—cognition by a person concerning that person's own cognitive processes—in human performance (e.g., Forrest-Pressley, MacKinnon & Waller, 1985; Weinert, 1987; Wellman, 1985, 1992). Gavelek and Raphael (1985) discuss a form of metacognition, called metacomprehension, which addresses the abilities of individuals to adjust their cognitive activity in order to promote more effective comprehension, in particular, the manner in which questions generated by sources external to the learner (i.e., from the teacher or text), as well as those questions generated by the learners themselves, serve to promote their comprehension of text. White and Gunstone (1989) argue that resolution of conflicting beliefs and permanent conceptual change requires "metalearning"—control over one's learning. For example, they discuss a study by Gauld (1986) that shows that students who learn new scientific beliefs often revert to their original beliefs over time because they have merely accepted the new knowledge without any real commitment to it. They argue that deep reflection on one's beliefs is a key part of the awareness and control over one's learning, and suggest methods for promoting metalearning in science classrooms.

It is clear, of course, that humans cannot exert explicit meta-control over all their learning processes, and the level of control that can be exerted, as well as how it is exerted, remain open questions. It is also possible (though, in our opinion, unlikely) that it may turn out not to be efficient to use this framework as a technological basis for the design of computer programs that learn. Nevertheless, the framework presented here may be used to take an intentional stance (Dennett, 1987) towards a reasoner for the purposes of building a computational model of learning. In such a stance, the competence of the reasoner can be modelled using goals, learning decisions, learning actions, and so forth as the basic theoretical constructs in task-level and algorithm-level descriptions of the reasoner. That stance can be taken without any commitment to existence of these constructs at the implementational level of, say, neural representations and processes in the human brain, or to the degree of conscious self-awareness of these processes in human thought.

4 A framework for goal-driven learning

In order to form a unified view of the diverse research results on goal-driven learning, we propose a general framework that describes the goal-driven learning process. While no single piece of research to date has investigated this framework as a whole or exactly as stated, the framework serves to provide an integrative structure into which individual research efforts fit as pieces of the puzzle of goal-driven learning. The key idea behind our framework is to model learning as an *active* (explicitly goal-driven) and *strategic* (rational and deliberative) process in which a *reasoner*, human or machine, explicitly identifies its goals in learning and attempts to learn by determining and pursuing appropriate learning actions via explicit reasoning about its goals, its abilities, and environmental opportunities.

In this framework, learning is motivated by the *performance tasks* that the reasoner is attempting to perform in the world. The performance tasks give rise to *task goals*, as well as subgoals of those goals, and subtasks to achieve them. As the tasks and subtasks are performed, the reasoner formulates explicit *learning goals* to perform types of learning which, if successful, would improve its ability to carry out those performance tasks or subtasks. The learning goals, in turn, guide the learning behavior of the reasoner, leading it to focus attention, allocate resources, and select appropriate *learning algorithms* or *learning strategies* when opportunities to learn arise. In our previous example, the top-level task goal would be to get a good buy on a stereo, which would spawn subtasks such as going to a store and purchasing the stereo. These subtasks give rise to learning goals to learn information needed to select the store and the stereo to buy. Some

Reasoner	An intelligent system, human or machine.
Performance task(s)	Overall task(s) that the reasoner is performing that create an effect on the external world.
Task goals	Specific goals and subgoals to be accomplished in order to accomplish the performance task.
Learning goals	Goals to learn (including learning by acquiring knowledge, reorganizing or reformulating knowledge, verifying hypotheses, etc.). These include both a description of the needed information and information about the task for which the information is needed.
Strategies or methods	Processing steps that accomplish a goal.
Algorithms	Computational formulations of strategies.
Reasoning trace	Record including information on goal-subgoal decomposition of goals, choice of methods to accomplish goals and subgoals, and other decisions taken in pursuing those goals, as well as the bases for these decisions, results of reasoning actions, alternative courses of action, etc.

Table 2: Summary of terminology.

of those learning goals may seek to gather information about the external world, while others may seek to create generalizations, test hypotheses, reorganize memory, or otherwise change existing knowledge. Those learning goals prompt the choice of learning strategies such as “shopping around,” looking at reviews in magazines, and so forth.

The goal-driven learning process involves not only learning about the world, but also learning to improve the reasoner’s own reasoning process. In order to identify the learning that needs to occur, the reasoner needs to be able to analyze its reasoning process in addition to the knowledge that the reasoner invokes during the reasoning process. To facilitate this, the reasoner maintains a *reasoning trace* of its internal decision-making. The reasoning trace provides the basis for *introspective reasoning* or *meta-reasoning* to guide learning and improve its reasoning performance. Table 2 summarizes the terminology we will use in our framework.

More concretely, goal-driven learning can be modeled as a two-step process. The first step involves the generation of learning goals based on the performance tasks and task goals of the reasoner. This step can be thought of as the process of deciding what to learn, and results in the formulation of learning goals that specify the desired learning that is to occur as well as the origin of the need for this learning. The second step involves the pursuit of learning goals based on the reasoner’s needs, its resources, and on environmental factors that determine the timeliness of pursuing certain learning actions in a given situation. This step can be thought of as the process of deciding how and when to learn and carrying out the learning. When the learning actually occurs, this step results in the satisfaction of one or more of the reasoner's learning goals.

Step 1: Generating learning goals: Figure 1 describes the process by which learning goals are generated. The reasoner is assumed to be pursuing a performance task that can be characterized in terms of the current situation and task goals specifying the desired result of the task. In the stereo example, the situation might be that the shopper lives in New York, knows nothing about stereos, and has \$500 to spend; the task goal would be to buy a stereo that was a good value for the \$500 price range.

Given the performance task, the reasoner performs reasoning in support of that task and maintains a trace reflecting its reasoning process. The reasoning trace records the goal-subgoal decompositional structure of the task goals, the choice of methods for achieving them and other decisions taken, the factors influencing those decisions, and descriptions of other reasoning actions (e.g., attempts to retrieve information) and their outcomes (Carbonell, 1986; Ram & Cox, 1994/chapter 7). For example, forming an executable plan to get a good buy on a stereo requires knowing which stereo to buy and where to buy it. If the reasoner does not know, a reasoning failure occurs because current knowledge is insufficient to make a decision.

At a suitable point in processing, the reasoning trace and its results are evaluated in light of the reasoner's task goals. If any problems arose during processing, learning is needed to enable the reasoner to avoid similar problems in the future. In being driven by deficiencies in the reasoner's knowledge, the process for generating learning goals is in the spirit of impasse-driven or failure-driven learning (e.g., Chien, 1989; Collins & Birnbaum, 1988; Hammond, 1989; Kocabas, 1994; Laird, Newell, & Rosenbloom, 1986; Mooney & Ourston, 1993; Mostow & Bhatnagar, 1987; Newell, 1990; Owens, 1991; Park & Wilkins, 1990; Ram & Cox, 1994/chapter 7; Riesbeck, 1981; Schank, 1982; Schank & Leake, 1989; Sussman, 1975; VanLehn, 1991a). There are several kinds of failures that may be involved, for example *expectation failures*, *retrieval failures*, or *knowledge application failures*. Expectation failures arise when the achieved outcome conflicts with expectations, regardless of whether the outcome was desirable (e.g., Collins & Birnbaum, 1988; Freed & Collins, 1993; Hammond, 1989; Leake, 1992; Owens, 1991; Ram & Cox, 1994/chapter 7; Schank, 1986). In our framework, an unexpected success is also treated as an expectation "failure." An example of a retrieval failure is the failure of a schema-based understanding program to retrieve an applicable schema, even though that schema exists in memory (Ram, 1993).

Knowledge application failures arise when retrieved knowledge structures fail to apply fully to new situations, and trigger learning to reconcile the conflicts (e.g., Kass & Leake, 1988; Leake, 1992; Mooney & Ourston, 1993; Park &

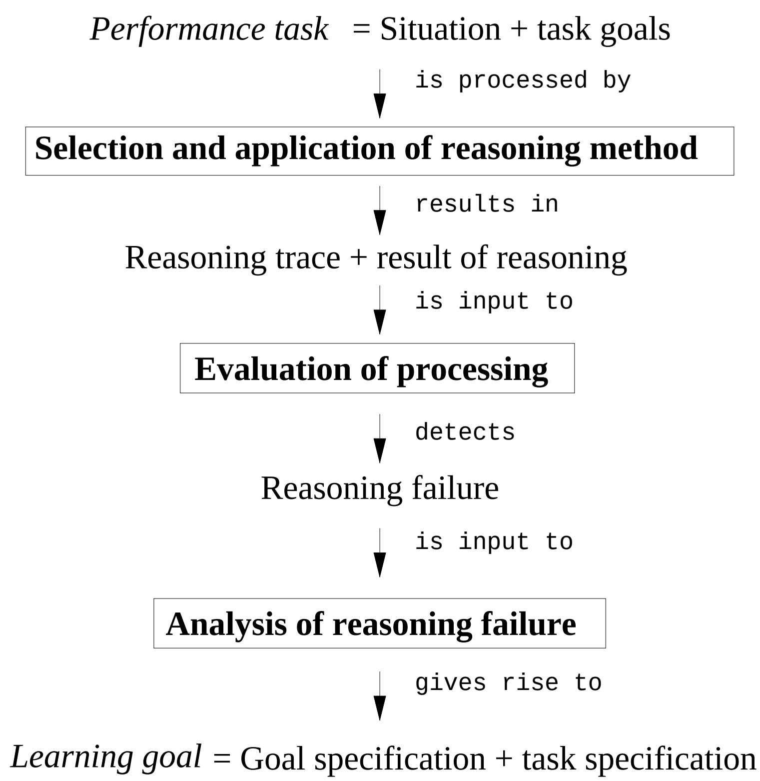

Figure 1: Generation of learning goals

Wilkins, 1990; Ram, 1993; Schank & Leake 1989). Ram, Cox, and Narayanan (chapter 18) present a taxonomy of possible types of failures and discuss their relationship to goal-driven learning.

Even if no failure has yet occurred, anticipation of a reasoning failure may trigger learning. For example, a reasoner may realize that it cannot perform a task and decide to perform the necessary learning before even attempting the task. In our framework, all these motivations for learning—reasoning failures, difficulties, impasses, suboptimalities, surprises, and other types of processing problems or anticipated processing problems—will be collectively and simply referred to as *failures*.

Different kinds of failures give rise to different kinds of learning goals. For example, a reasoner may need to acquire additional knowledge if its reasoning reached an impasse due to missing knowledge, as in the case of a novice stereo buyer who has no knowledge of which brand of stereo to buy. If the reasoner possessed sufficient knowledge but did not retrieve it at an appropriate time, it may need to reorganize memory (Ram & Cox, 1994/chapter 7; Ram, Cox, & Narayanan/chapter 18). A reasoner may need to modify the underlying representational vocabulary if its vocabulary is found to be inadequate (e.g., Schlimmer, 1987; Wrobel, 1988). In some situations, a reasoner might also need to add to its repertoire of reasoning strategies (e.g., Leake, 1993).

When a reasoning failure is detected, the reasoning trace is analyzed, in a process called credit/blame assignment, to find the source of the failure (Birnbaum, Collins, Freed & Krulwich, 1990; Hammond, 1989; Minsky, 1963; Ram & Cox, 1994/chapter 7; Weintraub, 1991). Blame assignment may be thought of as a process of model-based diagnosis of the reasoner itself (Birnbaum, Collins, Freed & Krulwich, 1990; Stroulia, Shankar, Goel, & Penberthy, 1992). If the failure is attributed to faulty knowledge, learning is needed to improve the reasoner's performance, and a learning goal is generated to repair that knowledge. In our framework, the learning goal is characterized in terms of two pieces of information: The desired learning—*what* learning is needed—and a description of the task that motivates learning—*why* learning is needed. The additional information about why learning is needed is important to allow the reasoner to carry out its tasks in an opportunistic manner, with learning goals (and the tasks that they support) being suspended until circumstances are favorable to their pursuit (Ram, 1989, 1991, 1993; Ram & Hunter, 1992/chapter 4).

Step 2: Pursuing learning goals: In the goal-driven view of learning, learning goals are treated analogously to task goals in the world. Just as task goals are achieved through a planning process using available methods for reasoning and action, learning goals are achieved through a *knowledge planning* process using available learning methods or strategies (Hunter, 1990/chapter 2; Quilici, in press; Ram & Hunter, 1992/chapter 4; Redmond, 1992). In the knowledge planning process, explicit reasoning is done about learning goals, their relative priorities, and strategies by which they can be achieved. These learning goals, also called *knowledge goals* (Ram, 1987, 1990; Ram & Hunter, 1992/chapter 4), can be represented in a goal dependency network (Michalski, 1993; Michalski & Ram/chapter 21), which is used to select and combine learning actions into learning strategies that are appropriate for current learning goals and for the learning opportunities provided by the current environment.

Individual learning actions may include performing knowledge acquisition (e.g., asking a friend to recommend a stereo) knowledge reorganization (e.g., grouping stores by the size of their stereo departments), knowledge reformulation or transmutation (e.g., forming new generalizations from stored episodes concerning others' experiences with particular sound equipment), and so on. Their application is guided by the learning goals of the reasoner (Gratch, DeJong, & Chien, 1994; Hunter, 1990/chapter 2; Michalski & Ram/chapter 21; Pryor & Collins, 1992/chapter 10; Ram & Cox, 1994/chapter 7; Ram & Hunter, 1992/chapter 4). Figure 2 sketches the second step of the goal-driven learning process. This step begins with reasoning about the relationships and relative priorities of learning goals in order to form a goal dependency network. Based on the information contained in the goal dependency network and on environmental factors affecting the appropriateness of different goals, the reasoner selects the learning goals to pursue. Learning strategies for achieving those goals in the current environment are then selected and applied.

Perspective on the framework: The model of learning embodied in the above steps contrasts with the approach to learning taken in traditional machine learning systems in artificial intelligence. Typically, in those systems, learning is primarily a passive, data-driven process of applying a single learning algorithm (or a predetermined combination of a few learning algorithms) to training examples presented to the system. Goal-driven learning, in contrast, is an active and strategic process driven by reasoning about information needs, alternative learning strategies, and opportunities in the environment. In our framework, the process of determining what to learn is an integral part of the computational model of learning, as is the process of deciding (on a dynamic basis) how and when to learn it.

Our view of goal-driven learning implies a tightly coupled relationship between learning and the "rest of reasoning." This view is consistent with recent models of intelligence that are framed as *integrated intelligent architectures* (sometimes known as *embedded systems*), in which the knowledge and reasoning tasks underlying learning and performance are integrated into a complete interacting system. Numerous approaches to such architectures were presented, for example, at the 1991 AAAI Spring Symposium on Integrated Intelligent Architectures (Laird, 1991), the 22nd Carnegie-Mellon Symposium on Cognition (VanLehn, 1991b) and at the Integrated Learning Workshop at the 1993 European Conference on Machine Learning (see Plaza, Aamodt, Ram, van de Velde,

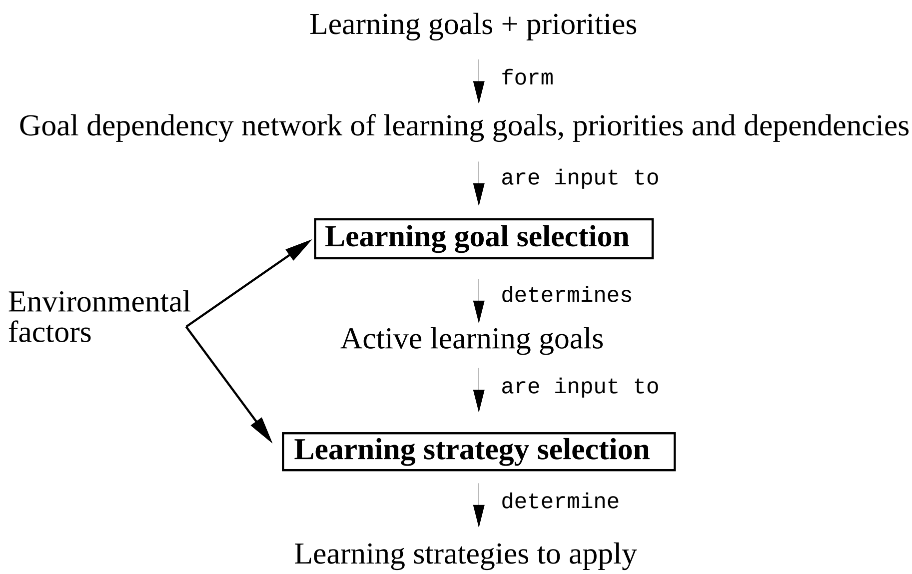

Figure 2: Pursuing learning goals using appropriate learning strategies

& van Someren, 1993, for an overview). Some of these architectures propose that one or a few primitive mechanisms underly intelligence (e.g., Soar (Rosenbloom, Laird, & Newell, 1993), ACT* (Andersen, 1983)), while others integrate many (often higher-level) mechanisms that cooperate to achieve the agent's task and reasoning goals (e.g., PRODIGY (Carbonell, Etzioni, Gil, Joseph, Knoblock, Minton, & Veloso, 1991/chapter 11) and Theo (Mitchell, Allen, Chalasani, Cheng, Etzioni, Ringuette & Schlimmer, 1991)). A common theme in this research, and one that is compatible with the goal-driven learning framework proposed here, is the explicit representation of task goals, reasoning goals, and learning goals, and their role in a multistrategy reasoning process that integrates learning with performance tasks such as problem solving or comprehension (e.g., see Ram, Cox, & Narayanan/chapter 18). This theme is also shared with recent approaches that focus more on knowledge and knowledge-intensive reasoning than on the underlying cognitive architectures, such as Aamodt's (1991) CREEK and Hinrich and Kolodner's JULIA (Hinrichs, 1992).

5 Major issues in goal-driven learning

Our framework suggests several issues that must be addressed by a general theory of goal-driven learning:

- **What is a goal?**

  The term "goal" has been used to refer to several theoretical constructs in previous theories of learning and reasoning, including tasks, problem solving outcomes, desired states of the world, target concepts for learning, policies and orientations for learning, and so on. Consequently, characterizing goal-driven learning depends on determining the meaning of "goals" as they apply to the learning process.

- **What are the types of goals?**

  Given the wide range of goals and other influences described in learning research, another central issue is to identify different types of goals, how they relate to one another, and how different formulations of goal-based influences on learning can be placed in a unified framework.

- **How do goals influence processing and learning?**

  A premise of the goal-driven learning framework is that reasoning about goals directs the learning process. A fundamental question is what effects goals actually have on the learning process and how their influence is achieved.

- **What are the functional and pragmatic implications of goal-driven learning for the reasoner?**

  Given the differences between goal-driven learning and traditional learning models, one key question concerns the effects of goal-driven learning: What are the functional implications of goal-driven learning for the reasoner's own performance? What are the pragmatic implications of goal-driven learning as a model of reasoning, and for the design of intelligent systems?

- **What are the pragmatic implications of goal-driven learning as a cognitive model?**

  Considering goal-driven learning as a cognitive model raises questions about the pragmatic implications of that model. One key question concerns the predictions that the model suggests, which have implications for testing and validating theories of goal-driven learning. Another concerns the implications of the model for practical applications such as the design of instructional material and educational environments.

The following sections start with the first three of these points, illustrating relevant distinctions using examples from artificial intelligence research, and then take a broader view in considering the pragmatic ramifications of the goal-driven learning model. In Leake and Ram (1993/chapter 16), we return to these issues, discussing the individual perspectives on them that were advanced by the panelists at the Symposium on Goal-Driven Learning held at the 1992 Conference of the Cognitive Science Society.

6 What is a goal?

As Barsalou (chapter 17) observes, in some sense any reasoner executing a built-in procedure can be viewed as having a "goal" to perform that type of processing, so that any learner could be considered trivially "goal-driven". To distinguish between built-in behaviors and behaviors that are more explicitly goal-driven, Barsalou differentiates between implicit background orientations and explicit problem solving or task goals. Explicit task goals are the goals that guide a problem solving process in which a person intends to achieve a set of goals, assesses what must be performed to achieve them, and executes the needed actions. In contrast, an implicit background orientation is a behavior that is performed without explicit reasoning about when and how it should be pursued. For example, one such implicit orientation is the orientation to constantly maintain a world model that adequately represents the reasoner's environment (e.g., Barsalou/chapter 17, Leake, 1989, 1992), although in some formalisms this is expressed in terms of an explicit goal (see, e.g., Van de Velde, 1988).

Explicit goals are traditionally expressed as specifications of a target or desired outcome of a problem solving or learning task (e.g., Fikes, Hart, & Nilsson, 1972; Newell & Simon, 1972). However, Ram and Hunter (1992/chapter 4; Hunter, 1990/chapter 2; Ram & Cox, 1994/chapter 7) argue that capturing the introspective nature of the goal-driven learning process requires a richer characterization in which a goal is not merely a specification of a target. They argue that a target specification or an orientation is a goal only if the reasoner can actively plan to accomplish the goal, can make decisions about it, and can even decide to suspend it or not to pursue it.

In order for the reasoner to make such decisions, goals must be explicitly represented, and the reasoner must be able to reflect on its goals, how to achieve them, and their relative priorities and interdependencies. Ram and Hunter (1992/chapter 4) discuss a representation of learning goals in terms of the desired knowledge to be learned as well as the reason that the knowledge is needed. Additional representational issues concern the kinds of decision-making relationships that goals can enter into (Thagard & Millgram/chapter 19) and the intergoal relationships and interdependencies in which goals can play a role (Cox & Ram, 1994; Michalski & Ram/chapter 21; Schank and Abelson, 1977; Slade, 1993; Wilensky 1983).

7 Types of goals

In order to understand how goals can relate to one another and to learning, it is useful to consider the classes of goals that influence learning processes in existing computational models of learning (implemented as computer systems that learn). This section examines those classes. Our taxonomy is divided into the classes centered around *task goals*, *learning goals*, and *specifications, policies, and constraints*. Broadly, task goals determine *why* the reasoner is learning in the first place, learning goals specify *what* the reasoner needs to learn, and specification, policies and constraints influence *how* learning occurs.[1]

7.1 Task goals

In many systems, goals are modelled as descriptions of desired results or states in an external world, which we call "task goals." Task goals, exemplified by early planning programs such as STRIPS (Fikes, Hart, & Nilsson, 1972) and NOAH (Sacerdoti, 1977), are specifications of desired outcomes from a performance task in the external world, which are explicitly pursued through planful reasoning processes or, in some recent models, goal-directed reactive processes (e.g., Earl & Firby, 1994; Firby, 1987; Freed & Collins, 1994; Maes, 1990).

In order to pursue its task goals, a system may need to reason about the task goals and reasoning goals of other agents. In order to understand a story (or a real-world situation) involving other intelligent agents, a computer understanding system needs to model the goals and plans of those agents (Schank & Abelson, 1977; Wilensky, 1978); in addition, consideration of such goals affects the comprehension process of humans (Abbott & Black, 1986). Reasoning about other actors' goals also plays an important role in AI models of subjective understanding (Carbonell, 1983; Ram, 1990). Representation of goals and goal interactions is central in understanding as well as in planning (Wilensky, 1983). Such representations must capture both task goals and reasoning goals of both other agents and the system itself.

Because task goals characterize a desired state of affairs, they can also be used to describe the need for information that a planner requires to achieve that state of affairs (e.g., Etzioni, Hanks, Weld, Draper, Lesh, & Williamson, 1992; Leake, 1991b/chapter 9; Ram & Leake, 1991), to understand interactions between task goals (e.g., Freed & Collins, 1994), and to influence or bias learning strategies (e.g., Martin, 1994; Provost, 1994). In some models, task goals (and resulting learning goals) are decomposed into subgoals or task structures to facilitate planning and learning (e.g., Karlsson, 1994; Stroulia & Goel, 1994). In planning systems that store prior plans, such as Hammond's (1989) CHEF, task goals drive the search for relevant plans in memory and trigger learning of new plans and new indices for plan retrieval when failures arise. A similar role is played by "problem goals" in Veloso and Carbonell's (1993) model of case-based reasoning in the PRODIGY system, and by functional specifications or "design goals" in design programs (e.g., JULIA, Hinrichs, 1992; Kolodner, 1987).

[1] To clarify the differences and commonalities in different approaches, in this discussion we will use a common vocabulary and framework to discuss individual pieces of research in an attempt to present an integrated view of goal-driven learning, even at the expense of sometimes differing from the terminology used by the original researchers.

7.2 Learning goals

Other computational models explicitly describe goals for learning, rather than implicitly characterizing it in terms of the external task. These learning goals differ from task goals in that, while they too specify a desired state, the specified state is an internal or mental state—a state of knowledge or belief that the learner is attempting to achieve. Task goals are satisfied through problem solving in the external (usually physical) world, while learning goals are satisfied through a learning process that, in the goal-driven learning framework, is viewed as problem solving in the "informational" world. These learning goals have been characterized in the following ways:

- **Knowledge acquisition goals or knowledge goals:** Schank and Abelson (1977) describe a category of knowledge goals (called "D-KNOW" goals) to determine needed information. In their model, such goals arise when a planner requires knowledge of particular facts (e.g., the location of a desired object) to achieve its other goals. The planner generates plans for satisfying these goals using standard methods for seeking information in the external world.

  The term "knowledge goal" was introduced by Ram (1987), and general knowledge goals are discussed in more detail in Ram & Hunter (1992/chapter 4). Ram (1990) proposes the use of knowledge goals as the basis for focus of attention in understanding and learning. Hunter's (1990/chapter 2) IVY and INVESTIGATOR programs identify and pursue "knowledge acquisition goals" whose satisfaction constitutes learning in those systems. Ram and Cox's (1994/chapter 7) Meta-AQUA system uses knowledge goals such as knowledge refinement goals, knowledge reconciliation goals, and knowledge differentiation goals, to specify desired learning in a multistrategy learning system. The system then reasons about and selects the learning algorithms most appropriate for achieving its knowledge goals (Cox & Ram, 1994). Knowledge acquisition goals in desJardins's PAGODA system (called "learning goals" in that system) represent concepts which, if learned, would maximize the system's expected utility (DesJardins, 1992/chapter 8).

- **Questions:** Ram's (1991, 1993) AQUA asks "questions" which are then represented as "knowledge goals." As in IVY, AQUA's learning occurs through satisfying knowledge goals, but using different methods; IVY looks for desired information in diagnostic cases, and AQUA tries to answer its questions by reading stories (Ram & Hunter, 1992/chapter 4). In Oehlmann, Sleeman, and Edwards' IULIAN (1992), questions and experimentation interact in an exploratory discovery process applied to the domain of electrical circuits.

- **Learning goals:** Michalski's (1993) MTL framework uses "learning goals" as the starting point for learning; relationships between "learning goals" are then used to combine basic knowledge transmutations into learning actions and to prioritize learning activities. Such learning goals subsume knowledge acquisition goals, knowledge organization goals to reorganize existing knowledge, as in AQUA, Meta-AQUA, and IVY, and knowledge reformulation goals as in Meta-AQUA and MTL.

In our further discussion we will consider that learning goals, in addition to specifying the desired outcome of learning, specify the reason that the desired learning is required (for example, AQUA's "task specifications" specify the suspended reasoning task that is awaiting the knowledge to be learned).

7.3 Specifications, policies, and constraints

Numerous computer systems reflect other types of influences and constraints on learning that are goal-related. Although these are not properly "goals" in our sense, because they do not drive the learning process in an explicit manner, they may play an important role in influencing that process. One such influence is:

- **Goal concepts or target concepts:** Mitchell, Keller, and Kedar-Cabelli's (1986) EBG algorithm, implemented in several computer programs, takes as input a "target concept" or a "goal concept," in order to learn an operational description of that concept (see also Minton, 1990/chapter 3).

  "Target concepts" are similar to "learning goals" in that they specify the desired outcome of learning. However, in accordance with our earlier discussion of goals, target concepts are better viewed as specifications of the desired output of a learning strategy rather than an explicit goal to learn, unless they are pursued through active, strategic, or planful reasoning processes. In addition, target concepts (unlike learning goals) do not specify the motivation for learning.

The following categories are all ways to characterize the value of learning for the learner. In particular, they describe the policies under which the learning task should operate in order to better achieve the overarching learning goals, and describe relevant constraints on the processes that carry out the learning task:

- **Purposes:** Kedar-Cabelli's (1987) PURFORM uses "purposes" of artifacts, defined in terms of their role in enabling plans, to determine target concepts for learning. Leake's (1991b/chapter 9, 1992) ACCEPTER guides explanation evaluation in terms of tasks in the world which give rise to "purposes" to build particular types of explanations, which in turn provide the information needed to satisfy system learning goals.

- **Operationality criteria:** In explanation-based learning systems, "operationality criteria" (surveyed in Keller, 1988) characterize requirements for useful concept descriptions.

- **Preference criteria, inductive biases, and general policies and constraints:** Inductive learning systems such as PREDICTOR (Gordon & Perlis, 1989/chapter 13), LEX (Mitchell, 1982), and STABB (Utgoff, 1986) use an "inductive bias" to restrict the space of candidate hypotheses. Michalski's (1983) INDUCE method uses a lexicographic preference criterion to rank candidate hypotheses for generalization. Laird, Rosenbloom, and Newell's (1986) SOAR system incorporates a "policy" to learn from each subgoal during problem solving.

- **Utility metrics:** Minton's (1990/chapter 3) PRODIGY system uses a "utility metric" to determine whether a piece of knowledge is worth learning or storing. Gratch and DeJong's (1993) COMPOSER uses "expected utility" to characterize the quality of a reasoner with respect to a task, which increases with learning. desJardins's (1992/chapter 8) PAGODA computes the utility of plans and the costs of planning and learning to guide learning. PAGODA's "learning goals" represent concepts which, if learned, would maximize the system's expected utility.

Policies and constraints are not learning goals in the sense that the learner does not actively seek to satisfy them; instead, they influence the learning processes that the learner uses to achieve its learning goals. In particular, they describe the policies under which the learning task should operate in order to better achieve the overarching learning goals, and describe relevant constraints on the processes that carry out the learning task. Note that a learner might formulate explicit learning goals to learn these criteria. For example, a learner might formulate an explicit goal to learn appropriate biases for a given type of learning situation, and pursue an explicit learning agenda to learn such biases.

7.4 A unifying view

The underlying commonality among these constructs is that each reflects an intention to influence learning according to needs that are external to the learning process itself. However, quite different focuses are apparent in the formulations described in the previous sections. Consequently, developing a general theory of goal-driven learning depends on analyzing the relationships of these constructs and their role in reasoning and learning.

To relate the previous perspectives, we will establish a uniform vocabulary. In the following discussion, we will refer to the general class of *goals* to describe theoretical constructs that refer to mental entities reflecting desired states that are explicitly represented and actively pursued through a planful reasoning process.[2] *Task goals* will refer to goals which specify desired effects in the world external to the reasoner. *Learning goals* or *knowledge goals* will refer to goals which specify desired effects within the reasoner such as acquiring new knowledge or augmenting, reorganizing, or reformulating existing knowledge. Learning goals describe not only the desired processing outcome, but how the desired knowledge will be used when it is acquired. *Reasoning goals* will refer to more general internal goals to form conclusions or inferences through learning or other reasoning processes.

We will also refer to *target concepts* that specify a desired concept to be learned, but not necessarily learned through a goal-driven learning process; and to general *policies* or *orientations* that influence learning without being explicitly represented or available for manipulation by the reasoner's reasoning or learning process, including *constraints* on the formulation of hypotheses such as biases, operationality criteria, and utility metrics. Our vocabulary makes the following distinctions:

- **Task goals vs. reasoning goals:** Task goals are goals to be achieved in the world external to the reasoning system; reasoning goals are achieved within the reasoning mechanism of the system.

- **Reasoning goals vs. learning goals/knowledge goals:** Reasoning goals span the broad range of deliberative activities, including activities such as retrieval of relevant information, similarity assessment, etc. Learning goals/knowledge goals refer solely to goals to acquire or formulate particular types of knowledge.

- **Goals vs. policies/orientations/constraints:** Goals involve specifications of internal or external states to be actively planned for; policies specify background orientations that are implicit in the processes that achieve these states.

- **Learning goals vs. target concepts:** Learning goals are manipulated by an explicit, strategic planning process, while target concepts are specifications of desired results from a learning algorithm that uses the specification only to evaluate its results, rather than to guide the on-going learning process.

Table 3 summarizes these distinctions. Note, however, that these classes of goals can overlap and influence each other. Task goals have been used to guide learning and performance in several systems, and can also be used to formulate learning goals to acquire information necessary for a given task (Ram & Leake, 1991) or to come to a better understanding of the task itself (Freed & Collins,

[2] Note that this definition does not imply that goals or goal-driven processing must necessarily be conscious, nor that the reasoner must necessarily be able to report externally about this processing.

	Explicitly represented?	Range of effects (internal to reasoner or in external world)	Influences selection of solution algorithm?	Solution process	Effect on solution generation
Goals	Yes	Either	Yes	Planning	Guidance
Task goals	Yes	External	Yes	Planning actions in external world	Guidance
Reasoning goals/ learning goals	Yes	Internal	Yes	Knowledge planning	Guidance
Policies	Sometimes	Internal	Sometimes	Unspecified	Constraint
Target concepts	Yes	Internal	No	Unspecified	Guidance
Operationality criteria	Yes	Internal	No	Unspecified	Constraint

Table 3: Types of goals and policies.

1994). In conjunction with knowledge or theories, they can guide learning processes (Barsalou, 1991/chapter 5; Ng & Bereiter, 1991/chapter 14; Wisniewski & Medin, 1991/chapter 6). Likewise, although target concepts are generally provided to a learning system as input by a human user, in some models target concepts are generated from aspects of the performance task in a manner similar to the generation of learning goals. For example, Kedar-Cabelli (1987) discusses a method for generating target concepts from standard constraints on artifacts to be used in particular plans. Keller (1987) also sketches a process for generating learning goals from higher-level performance objectives. Similarly, policies (such as bias, which is usually formulated as a passive, background constraint on learning) may be actively monitored and modified by the reasoner to guide the learning task (Gordon & Perlis, 1989/chapter 13; Martin, 1994; Provost, 1994; Provost & Buchanan, 1992; Utgoff, 1986).

Several models include learning goals as an explicit part of their formulation of the learning process. Learning goals have been used to guide resource allocation, information search, hypothesis evaluation, and other aspects of learning; to select and combine learning strategies; to guide and to learn about the reasoning process itself; and to model active learning in educational contexts. These models are discussed in the following sections.

8 Role of goals in learning

Given the range of goals that can influence learning, it is not surprising that different models reflect different types and degrees of goal-based influence. Thagard and Millgram (chapter 19) propose a broad distinction between learning which is explicitly *goal-driven* and that which is *goal-relevant*. Goal-relevant processing is not explicitly directed by the goals of the reasoner, but results in outcomes that are nevertheless useful with respect to those goals. Thus goal-relevant processing is similar to Barsalou's (chapter 17) orientations in which the desired learning may occur as a side-effect of normal task-related processing. For example, a reasoner may have an implicit orientation to maintain an accurate model of the world around it (Barsalou, chapter 17; Leake, 1992). Goal-driven learning, in contrast, is driven by explicit learning goals of the reasoner; those goals influence or even determine the content of what is learned. As the reasoner's goals change, so does the learning that results.

One of the issues involved in goal-driven learning is how to balance competing goals to determine the goals and goal priorities that form the background for the goal-driven learning process (Thagard & Millgram, chapter 19). Once learning goals have been identified and prioritized, they can influence the system performance task, the system learning task, and the storage of results. The ways that goals can exert their influences are summarized in table 1, and are discussed in more detail in the following sections.

8.1 Guiding the performance task

In any goal-driven system, the influence of goals on the performance task also influences what is learned, by determining the focus of processing or changing the context in which learning is performed (Barsalou, chapter 17). For example, in case-based reasoning systems, the goals that drive processing also influence what is eventually learned (e.g., Kolodner, 1993; Hammond, 1989; Hinrichs, 1992; Ram, 1993; Veloso & Carbonell, 1993).

In addition, just as performance tasks can give rise to learning goals (e.g., the stereo buyer's learning goals), learning goals can themselves prompt and guide new performance tasks in service of the learning goals. For example, learning goals may guide tasks to gather needed information in the world, or to produce a situation in the world that is favorable to learning. Performance tasks may include reasoning tasks that are largely internal to the reasoner. For example, Leake (chapter 20) presents a model of explanation construction that is an integral part of a goal-driven reasoning and learning system. The reasoner can decide when explanations are needed, can characterize its information needs (goals), and can use this characterization to focus the search for explanations. Thus goals are used to guide the control procedure used in the performance task (explanation construction) and to manage the resources available for that task. Since explanation is a central part of learning, a goal-driven explainer is necessarily a goal-driven learner as well. Many performance tasks involve interaction with the outside world. For example, Ram (1991; Ram & Hunter, 1992/chapter 4) presents a model of natural language story understanding in which goal-driven processes are used to analyze and interpret natural language text (see also Carpenter & Alterman, 1994). Similarly, in Pryor and Collins's (1992/chapter 10) model, goals are used to guide the perception of visual images.

Xia and Yeung (1988/chapter 12) use goal-based considerations to learn new classifications of problem-solving strategies, and Hunter (1990/chapter 2) uses knowledge goals to guide the search for information by, for example, formulating appropriate queries to a database.

8.2 Guiding the learning task

The central tenet of goal-driven learning, and the thesis of this book, is that the learning is guided by explicit consideration of the reasoner's goals. The goal-driven learning framework involves first formulating explicit goals to find or infer certain beliefs, and then using these goals to drive reasoning and learning—which amounts to explicit decision-making and control. The resulting control of learning can be realized in several ways as described below.

8.2.1 Specifying the target of learning

Barsalou (1991/chapter 5) shows that people often derive categories in a dynamic, ad hoc manner during the construction of plans to achieve goals. For example, while foods are normally categorized into grains, vegetables, fruits, and so on, different category structures may be appropriate in the context of particular goals, giving rise to categories such as "foods to eat while on a diet". Some of these goal-derived categories become reasonably well-established for people or cultures in which the goal occurs often (such as dieting), but others remain ad hoc (such as "activities to do on a vacation in Japan with one's grandmother"). Thus learning can involve the construction of concepts that must be determined in a dynamic manner based on the demands of the particular task at hand. This is consistent with Ng and Bereiter's (1991/chapter 14) results on task- and goal-driven learning in an educational setting.

Wisniewski and Medin (1991/chapter 6) show that prior knowledge and intuitive theories can also influence learning (see also Murphy & Medin, 1985). They argue that tightly coupled interactions exist between knowledge and experience during learning. To the extent that learning is incremental, candidate hypotheses and theories learned earlier can influence later learning. In addition to previously learned theories, a goal-driven learner will also have previously formulated and possibly only partially satisfied learning goals as part of its learning context; these goals can also influence future learning (e.g., Ram, 1991, 1993).

In early artificial intelligence models of learning, goals were pre-specified as targets of particular learning algorithms. Such models did not have explicit learning goals; rather, they could be viewed as possessing background orientations to ensure that the learning actions are goal-relevant. However, task goals can be used to determine learning goals which specify the desired outcome of the learning task, whether it be a new piece of knowledge to be acquired or a new organization or formulation of existing knowledge (Ram & Leake, 1991). For example, Kedar-Cabelli (1987) and Keller (1987) propose extensions to earlier models in which the reasoner proposes its own targets rather than relying on an outside user to specify them.

In general, multiple learning goals are possible in any complex situation, and complex reasoning processes may be needed to determine which learning goals to generate (Krulwich, 1994; Ram, Cox, & Narayanan/chapter 18). It has been proposed that if learning is integrated with the reasoning process that it is in support of, an analysis of the reasoning process can be used to formulate learning goals (Hunter, 1990/chapter 2; Leake/chapter 20; Ram & Cox, 1994/chapter 7; Ram, Cox, & Narayanan/chapter 18; Ram & Hunter, 1992/chapter 4). For example, if a reasoner is unable to perform its task due to a missing piece of knowledge, it can formulate an explicit goal to learn that piece of knowledge. Learning goals may also seek to augment knowledge in other ways (e.g., learning a new antecedent to a rule (Mooney & Ourston, 1993; Park & Wilkins, 1990)), reorganize knowledge (e.g., learn a new index to an existing knowledge structure (Hammond, 1989; Ram, 1993)) or to reformulate existing knowledge (e.g., operationalization of abstract knowledge into a more directly usable form (Keller, 1988; Mostow, 1983); generalization or abstraction of examples (Michalski, 1993); modification of representational framework or vocabulary (Schlimmer, 1987; Wrobel, 1988)). Michalski (1993; Michalski & Ram/chapter 21) presents a taxonomy of the kinds of "knowledge transmutations" that may be used for various kinds of learning. As those learning goals are pursued, new learning goals may be generated on the fly, or existing learning goals abandoned, in response to changes in circumstances, the learner's knowledge, and overarching goals of the learner (Leake/chapter 20). Due to the dynamic nature of learning goals, the learning process itself must be dynamic as well.

8.2.2 Specifying the learning algorithms used

Many recent models of learning in humans and machines appeal to multiple methods for learning and reasoning in general (e.g., Michalski & Tecuci, 1993). For example, Ram, Narayanan, and Cox (1993) present a computational model of troubleshooting, based on a study of human troubleshooters on an electronics assembly line. In that model, several different learning strategies are used to improve troubleshooting performance, including compilation of knowledge through experience with specific problems, interactive acquisition of knowledge from a human expert, postponement of a problem until a later time, and forgetting knowledge that is made obsolete through changes in the manufacturing process.

In a wide range of artificial intelligence systems, goals determine the learning algorithms used to accomplish needed learning (e.g., Hunter, 1990/chapter 2; Michalski & Ram/chapter 21; Ram & Cox, 1994/chapter 7; Ram, Cox, & Narayanan, chapter 18; Ram & Hunter, 1992/chapter 4; Redmond, 1992; Quilici, in press; Xia & Yeung, 1988/chapter 12). A central issue in such models is how to select and combine appropriate learning algorithms in a given learning situation. This multistrategy learning process can be modeled with an explicit decision stage in which the appropriate learning strategy or strategies are identified (as shown in figure 2), followed by a strategy application stage in which the corresponding algorithm is executed (Ram & Cox, 1994/chapter 7). In addition, some models incorporate an explicit evaluatory phase in which the quality of a learned piece of knowledge is assessed. In those models, learning goals can be used to guide evaluation, such as in Leake's (1991b/chapter 9, chapter 20) use of goals to evaluate causal explanations.

8.2.3 Constraining the learning process

In some models of goal-driven learning, goals are used not to specify the desired target of learning or to select learning strategies, but rather to provide constraints to the process used for learning itself. For example, a reasoner that learns through inductive generalization must select from among a potentially very large number of possible hypotheses consistent with its inputs, and selecting the right candidate hypotheses can have an enormous effect on the ability of the reasoner to perform a particular task. Because the inputs do not adequately constrain the set of candidate hypotheses, hypothesis selection must be done via some criterion external to the inputs themselves. Such a criterion is called a "bias" (Mitchell, 1980) or "preference criterion" (Michalski, 1983). Although many early models of inductive learning appealed to a pre-determined bias, it is often advantageous for the learner to modify its bias (Utgoff, 1986). Gordon and Perlis (1989/chapter 13) discuss a computational model of inductive learning in which the formation of useful generalizations is facilitated by use of explicit biases; Provost and Buchanan (1992) discuss the use of "inductive policies" to automatically adjust the bias in a learning system based on explicitly represented tradeoffs; and Hadzikadic and Yun (1988) argue that concept formation should be viewed as a goal-driven, context-dependent process to assure its flexibility, efficiency, and generality.

Similarly, some computational models of analogical learning include a mechanism by which the reasoner's task goals can influence the analogical mapping process (e.g., Forbus & Oblinger, 1990; Kedar-Cabelli, 1987). As with other types of learning mechanisms, the functional justification for this influence is to ensure that the inferences made during learning are actually relevant to the overall performance task of the reasoner. Spellman and Holyoak (1993) present evidence for the influence of goals on analogical mapping in human learners as well.

8.2.4 Focusing search for information to carry out learning

Carrying out the desired learning may require acquisition of new information. Consequently, learning goals may give rise to processes that attempt to seek that information from the outside world. Some of the possible methods include reading text (Ram, 1991), querying a database (Hunter, 1990/chapter 2), active experimentation (Carbonell & Gil, 1990; Cohen, Kulikowski, & Berman, 1993; Rajamoney, 1993), or other planful activity to gather needed information (Etzioni, Hanks, Weld, Draper, Lesh, & Williamson, 1992; Leake/chapter 20; Pryor & Collins, chapter 10).

8.2.5 Determining when learning should be attempted

Having identified what to learn and how to learn it, a reasoner, in general, still needs to determine when (or whether) to perform the actions that will lead to the desired learning. This decision can be broken down into two fundamental aspects: identifying learning opportunities, and evaluating the potential utility of learning.

Identifying learning opportunities: It would be naive to expect the real world to be structured so as to facilitate the satisfaction of each individual's needs whenever they should arise. Instead, it is likely that goals (whether task goals or learning goals) will not be immediately satisfiable at the time when they are formulated. For example, a detective with the goal of identifying a criminal will usually need to do much investigation before having sufficient information to assign responsibility (Leake/chapter 20). Likewise, the reasoner may not have the resources to pursue all its goals all the time, forcing the reasoner to select particular goals to pursue (a detective performing multiple investigations will prioritize them according to their importance).

Furthermore, the real world environment may not provide the opportunity to pursue a particular goal even if the reasoner does decide to pursue it. For example, pursuit of a learning goal may require environmental resources (such as a library, or access to an expert) which may simply not be available at the time. In Ram's (1991) model of natural language story understanding, for example, the reasoner's questions about the story being read may not be answerable due to insufficient information being available in the story. Determining whether a suicide bombing is the work of a religious fanatic depends on having more information about the bomber than is likely to be available in the first accounts of the incident. Thus it is essential for the reasoner to be able to suspend its pursuit of a goal, and to be able to resume its learning processes when an opportunity to satisfy the goal presents itself (Ram & Hunter, 1992/chapter 4). An analogous argument has been made for the opportunistic pursuit of task or problem solving goals (Birnbaum & Collins, 1984; Hayes-Roth & Hayes-Roth, 1979; Hammond, Converse, Marks, & Seifert, 1993; Patalano, Seifert, & Hammond, 1993), and similar factors are relevant to learning behavior as well.

Evaluating the potential utility of learning: Not all learning is useful; learning may sometimes be undesirable if it leads to the accumulation of knowledge that is seldom used, that is not expressed in an efficient manner, or that impairs performance of the reasoner (Etzioni, 1992; Francis & Ram, 1993; Gratch & DeJong, 1993; Minton, 1990/chapter 3; Tambe, Newell, & Rosenbloom, 1990). Estimates of the potential utility of learned knowledge can be used to decide what to learn about, based on an analysis of the expected utility of the learning goals if they were to be achieved (e.g., desJardins, 1992/chapter 8). As mentioned earlier, while human learners may not be able to explicitly control their own reasoning processes to such a fine-grained level of detail, it is nevertheless possible to model human learning behavior in a goal-driven, utility-theoretic, or rational formalism (e.g., Anderson, 1991).

8.2.6 Evaluating the results of learning

The final criterion for the effectiveness of learning is how well the results of learning match the desired effects of the learning process. This question has received less attention than the questions of how goals influence initial learning. However, such evaluation has been used to determine whether to store the results of the learning algorithm (Minton, 1988; 1990/chapter 3), and could be used to formulate new learning goals in light of current results of the learning process.

8.3 Guiding storage

Evaluating the results of learning can a system to decide whether to store them, and, if it does, to decide how to store them. Minton (1988; 1990/chapter 3) proposes a learning process that estimates the usefulness of the generalized rules that it forms before they are actually stored in the rule library. The estimate is used to predict the utility of storing a new rule; only those rules expected to actually improve performance are stored. Once a rule has been stored, the effects of that rule on performance are monitored; rules that do not improve performance are removed from the rule library.

This kind of selective storage and retention of learned rules is an instance of a more general kind of goal-directed control of learning called *information filtering* (Markovitch & Scott, 1993). Information filters can be used to decide which learned items to store, which to retain in memory over time, which to apply in a given situation, and even which training experiences it should seek out and which it should learn from. The reasoner's task goals guide the filters in selecting what learning occurs and what products of learning are retained and used.

Despite the practical benefits of using goals to guide storage, claims of cognitive validity for such a model are controversial (see Barsalou/chapter 17 for arguments against direct goal-based guidance in human learners). However, regardless of whether explicit reasoning about goals affects individual storage decisions, goals can still have a profound effect on what results are stored by determining the course of processing. If goals determine processing and if processing determines storage, then goals determine storage indirectly. For example, in dynamic memory systems, processing intrinsically changes memory without the changes to memory necessarily being under explicit strategic control (Schank, 1982).

Goals also play a more direct role in storage in many AI systems. For example, case-based planning systems index learned plans according to the goals that those plans satisfy (e.g., Hammond, 1989; Hinrichs, 1992; Kolodner, 1987; Veloso & Carbonell, 1993). In this way, these systems organize their memories to facilitate re-use of those plans to accomplish similar goals. Likewise, case-based explanation systems index explanations according to the knowledge goals that the explanations were formulated to satisfy (e.g., Leake, 1991a, 1992; Ram, 1993; Schank & Leake, 1989). Thus, these systems attempt to store learned information to facilitate accomplishing similar future goals. Psychological experiments by Patalano, Seifert and Hammond (1993) also suggest that, in humans, goal-based factors can affect storage so as to facilitate noticing information relevant to the pursuit of those goals.

The preceding sections show that goal-driven learning provides flexibility for processing and the ability to tailor learning to current learner needs, helping to perform effective learning of useful information. Insofar as learning is simply a kind of reasoning (Ram, Cox, & Narayanan, chapter 18), many of the mechanisms of attention focussing and goal-driven processing in learning and in other reasoning will be identical. In particular, many of the results from research in planning may, mutatis mutandis, carry over to goal-driven learning as well.

9 Pragmatic implications of goal-driven learning

The previous discussion shows how goal-driven learning can provide considerable power in intelligent systems, whether those systems are viewed as computational models of human intelligence, or purely as artificial intelligence systems. In learning systems, goals can be used to focus learning and to avoid unrestricted search and inferencing. They can also be used to guide the information-seeking process and to make decisions about what, when, and how to learn.

Applying a planful model of learning promises to be fruitful for many applications, including perception (Pryor & Collins, 1992 10), intelligent information retrieval (Ram & Hunter, 1991), learning through apprenticeship (Redmond, 1992), knowledge acquisition (Quilici, in press), information search during explanation (Leake, chapter 20), medical diagnosis (Hunter, 1990), natural language understanding (Carpenter & Alterman, 1994; Cox & Ram, 1994; Ram, 1991), and manufacturing (Perez, 1994; Ram, Narayanan, & Cox, 1993).

In addition, goals can be used as a theoretical device to build computational models of strategic and active reasoning and learning processes, and such models have practical ramifications for the design of instructional material. Ng and Bereiter (1991/chapter 14) show that different kinds of goals facilitate different kinds of reasoning and result in different kinds of learning. Such results suggest principles for the design of computer-based tools for education (Scardamalia & Bereiter, 1991). For example, van Berkum, Hijne, de Jong, van Joolingen, and Njoo (1991/chapter 15) use goal-driven learning both as a theoretical framework for decomposing the education problem and as a guide toward designing simulation-based instructional software. Schank proposes that because of the importance of goals in motivating and guiding learning, instruction should be conducted using a particular type of simulation environment—a *goal-based scenario*—to exploit the role of learning goals (Schank, Fano, Jona, & Bell, 1993/1994). In goal-based scenarios, students play roles that are connected to their goals, and whose successful completion requires acquisition of the skills to be taught. In that way, goal-based scenarios provide a framework for students to perform goal-driven learning to acquire the skills to be taught.

To design educational environments that facilitate learning, one must understand the goal-driven information-seeking processes of the students who will be interacting with the environment, in order to encourage development of those processes. For example, in Scardamalia and Bereiter's (1991) Teacher C model, the teacher is concerned with helping students formulate their own goals, do their own activation of prior knowledge, ask their own questions, direct their own inquiry, and do their own monitoring of comprehension. Ng and Bereiter (1991/chapter 14) identify three types of goal orientation in learning: task completion, instructional, and knowledge-building. They show that students with knowledge-building goal orientation learn better—those students actively construct learning agendas, use prior knowledge in learning, and use the new learning to reconsider their prior knowledge.

Cognitive science research has shown that people learn by interpreting and constructing information (e.g., Resnick, 1983, 1987). Learning is viewed as a constructive, knowledge-building process (and often a collaborative one) rather than one of absorption (Bereiter, 1994; Roschelle, 1992). This principle has been used as the basis for the design of instructional scenarios which facilitate goal-driven interpretation and construction of knowledge (e.g., Edelson, 1993; Ng & Bereiter, 1991; Schank, Fano, Jona, & Bell, 1993/1994). Van Berkum, Hijne, de Jong, van Joolingen, and Njoo (1991/chapter 15) discuss learning environments in which computer simulations are used for instruction. They too distinguish between the learner's goals and the instructor's goals, and identify four aspects of the design of such systems: simulation models, learning goals, learning processes, and learning activity. In their model, learning goals have three dimensions: knowledge category (type of knowledge), knowledge representation (representation of that knowledge), and knowledge scope (generality and applicability of that knowledge). Learning occurs through interaction with simulated environments using four types of learning actions (orientation, hypothesis generation, testing, and evaluation) which are guided by the learning goals. As in Ng's model, learning is modelled as an active, constructive, and exploratory process, and the educational environment is designed to support such a process.

To this point, such applications of goal-driven learning models have been pursued independently of investigations of computer models of the goal-driven learning process itself. However, one of the goals of this volume is to show—by bringing together perspectives from diverse communities—the contributions that results from divergent research perspectives can make to one another, and to highlight the common questions that remain to be addressed. The previous discussion suggests the value of analyzing the goal-driven learning processes in humans as the basis for the design of instructional material and educational environments. In addition to its obvious practical value, this approach can be useful in empirically validating theories of goal-driven learning.

10 Summary

Computational and psychological investigations of goal-driven learning have addressed, broadly speaking, issues of what to learn, whether to learn, how to learn, and when to learn. In goal-driven learning, decisions about what, whether, how, and when to learn are determined by explicit reasoning about the reasoner's needs for information. Although many aspects of goal-driven learning have been investigated in diverse fields, that research has been conducted in a piecemeal fashion, largely segregated by field. Even when multiple studies have been conducted in a single field, as is the case for artificial intelligence, each study has tended to concentrate on a few aspects of the problem without placing those aspects within a unifying framework and examining their larger implications.

This chapter has presented a unifying picture of existing goal-driven learning research in terms of a new framework for modeling goal-driven learning, in terms of the types of goals that may guide learning, and in terms of the ways those goals can influence learning. The chapter has also discussed some of the pragmatic ramifications of the goal-driven learning model, both for intelligent systems and for educational applications. Those ramifications provide motivations for advancing our understanding of the goal-driven learning process.

The framework presented here is not suggested as a final theory of goal-driven learning, but rather a device for understanding the relationships of different results relevant to goal-driven learning and for suggesting issues that must be addressed with further investigation through a coordinated multidisciplinary research effort. The individual models and perspectives of the following chapters illuminate specific aspects of the framework and the issues that remain to be addressed in future research.

Acknowledgements

This research was supported in part by the National Science Foundation under grants IRI-9009710 and IRI-9409348, by the Air Force Office of Sponsored Research under contract F49620-94-1-0092, by the Georgia Institute of Technology, and by Indiana University. We would like to thank Betty Stanton for her enthusiasm and support for our work on this volume, and to thank the MIT Press anonymous reviewers for their detailed comments on this and the other chapters in this volume, as well as on the composition of the volume itself. We would also like to thank Michael Cox, Eric Domeshek, Susan Fox, Anthony Francis, Janet Kolodner, Kenny Moorman, Paul Panaro, and Raja Sooriamurthi for their many helpful comments on earlier drafts of this chapter.

References

Aamodt, A. (1991). *A Knowledge Intensive Approach to Problem-Solving and Sustained Learning* . Ph.D. Dissertation. University of Trondheim, Norwegian Institute of Technology, May, 1991.

Abbot, V., & Black, J. (1986). Goal-Related Inferences in Comprehension. In J. Galambos, R. Abelson, & J. Black, editors, *Knowledge Structures*, pages 123-142, Lawrence Erlbaum Associates, Hillsdale, NJ.

Anderson, J. (1983). *The Architecture of Cognition*, Harvard University Press, Cambridge, MA.

Anderson, J. (1991). The Place of Cognitive Architectures in a Rational Analysis. In K. VanLehn, editor, *Architectures for Cognition*, pages 1–24, Lawrence Erlbaum Associates, Hillsdale, NJ.

Barsalou, L. (1991/chapter 5 of this volume). Deriving Categories to Achieve Goals. In G.H. Bower, editor, *The Psychology of Learning and Motivation: Advances in Research and Theory*, Volume 27, Academic Press, New York, NY.

Barsalou (chapter 17 of this volume). Storage Side Effects: Studying Processing to Understand Learning. In A. Ram & D.B. Leake, editors, *Goal-Driven Learning*, MIT Press/Bradford Books, Cambridge, MA.

Bereiter, C. (1994). Paper presented at the AAAI Spring Symposium on Goal-Driven Learning, Stanford, CA.

Birnbaum, L. & Collins, G. (1984). Opportunistic Planning and Freudian Slips. In *Proceedings of the Sixth Annual Conference of the Cognitive Science Society*, pages 124–127, Boulder, CO.

Birnbaum, L., Collins, G., Freed, M., & Krulwich, B. (1990). Model-Based Diagnosis of Planning Failures. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 318–323, Boston, MA.

Carbonell, J.G. (1986). Derivational Analogy: A Theory of Reconstructive Problem Solving and Expertise Acquisition. In R.S. Michalski, J.G. Carbonell, & T.M. Mitchell, editors, *Machine Learning II: An Artificial Intelligence Approach*, pages 371–392, Morgan Kaufman Publishers, San Mateo, CA.

Carbonell, J.G., Etzioni, O., Gil, Y., Joseph, R., Knoblock, C., Minton, S., & Veloso, M. (1991/chapter 11). Planning and Learning in PRODIGY: Overview of an Integrated Architecture. *SIGART Bulletin*, 2(4).

Carbonell, J.G. & Gil, Y. (1990) Learning by Experimentation: The Operator Refinement Method. In Y. Kodratoff and R. Michalski, editors, *Machine Learning III: An Artificial Intelligence Approach*, pages 191–213, Morgan Kaufman Publishers, San Mateo, CA.

Carpenter, T. & Alterman, R. (1994). *A Reading Agent.* In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 63–67.

Chien, S. (1989). Using and Refining Simplifications: Explanation-based Learning of Plans in Intractable Domains. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 590-595, Detroit, MI.

Cohen, D., Kulikowski, C., & Berman, H. (1993). Knowledge-Based Generation of Machine Learning Experiments: Learning with DNA Crystallography Data. In *Proceedings of the Intelligent Systems for Molecular Biology Symposium*, pages 92–100.

Collins, G. & Birnbaum, L. (1988). An Explanation-Based Approach to the Transfer of Planning Knowledge Across Domains. In *Proceedings of the 1988 AAAI Spring Symposium on Explanation-based Learning*, pages 107–111, Stanford, CA.

Collins, G., Birnbaum, L., Krulwich, B. & Freed, M. (1991). Plan Debugging in an Intentional System. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, pages 353-358, Sydney, Australia.

Cox, M.T. (1993). *Introspective Multistrategy Learning.* Technical Report GIT-CS-93/02, Cognitive Science Program, College of Computing, Georgia Institute of Technology, Atlanta, GA.

Cox, M.T. & Ram, A. (1994). Managing Learning Goals in Strategy Selection Problems. In *Proceedings of the Second European Workshop on Case-Based Reasoning*, pages 85–93, Chantilly, France.

Dennett, D. (1987). *The Intentional Stance.* Bradford Books/MIT Press, Boston, MA.

desJardins, M. (1992/chapter 8 of this volume). Goal-Directed Learning: A Decision-Theoretic Model for Deciding What to Learn Next. In *Proceedings of the ML-92 Workshop on Machine Discovery*, pages 147–151, Ninth International Machine Learning Conference, University of Aberdeen, Scotland.

Earl, C. & Firby, R.J. (1994). An Integrated Action and Learning System. In M. desJardins & A. Ram, editors, *Proceedings of the AAAI Spring Symposium on Goal-Driven Learning*, pages 22–27, Stanford, CA.

Edelson, D.C. (1993). *Learning from Stories: Indexing and Reminding in a Socratic Case-Based Teaching System for Elementary School Biology.* Ph.D. dissertation, Technical Report #43, Northwestern University, Institute of the Learning Sciences, Chicago, IL.

Etzioni, O. (1992). An Asymptotic Analysis of Speedup Learning. In *Machine Learning: Proceedings of the Ninth International Workshop*, pages 129–136, 1992.

Etzioni, O., Hanks, S., Weld, D., Draper, D., Lesh, N., & Williamson, M. (1992). An Approach to Planning with Incomplete Information. In *Proceedings of the International Conference on Knowledge Representation*, Morgan Kaufmann, San Francisco, CA.

Faries, J. & Reiser, B. (1988). Access and Use of Previous Solutions in a Problem Solving Situation. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, pages 433–439, Montreal.

Firby, J. (1987). Adaptive Execution in Complex Dynamic Worlds. Ph.D. Dissertation, Research Report #672, Yale University, Department of Computer Science, New Haven, CT.

Fikes, R.E., Hart, P.E., & Nilsson, N.J. (1972). Learning and Executing Generalized Robot Plans. *Artificial Intelligence*, 3:251–288.

Forbus, K.D. & Oblinger, D. (1990). Making SME Greedy and Pragmatic. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, pages 61–68, Cambridge, MA.

Forrest-Pressley, D.L., MacKinnon, G.E., & Waller, T.G. (1985), editors. *Metacognition, Cognition, and Human Performance, Volume 1 (Theoretical Perspectives)*, Academic Press, New York.

Francis, A. & Ram, A. (1993). Computational Models of the Utility Problem and their Application to a Utility Analysis of Case-Based Reasoning. In *Proceedings of the ML-93 Workshop on Knowledge Compilation and Speedup Learning*, Tenth International Machine Learning Conference, University of Massachusetts, Amherst, MA.

Freed, M. & Collins, G. (1993). A Model-Based Approach to Learning from Attention-Focusing Failures. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, pages 434–439, Boulder, CO.

Freed, M. & Collins, G. (1994). Learning to Cope with Task Interactions. In M. desJardins & A. Ram, editors, *Proceedings of the AAAI Spring Symposium on Goal-Driven Learning*, pages 28–35, Stanford, CA.

Gauld, C. (1986). Models, meters and memory. *Research in Science Education*, 16:49–54.

Gavelek, J.R. & Raphael, R.E. (1985). Metacognition, Instruction and the Role of Questioning Activities. In D.L. Forrest-Pressley, G.E. MacKinnon, & T.G. Waller, editors, *Metacognition, Cognition, and Human Performance, Volume 2 (Instructional Practices)*, pages 103–136, Academic Press, New York.

Gordon, D. & Perlis, D. (1989/chapter 13). Explicitly Biased Generalization. *Computational Intelligence*, 5(2):67-81.

Graesser, A.C., Person, N., & Huber, J. (1992). Mechanisms that Generate Questions. In T.W. Lauer, E. Peacock, and A.C. Graesser, editors, *Questions and Information Systems*, pages 167–187, Lawrence Erlbaum Associates, Hillsdale, NJ.

Gratch, J. & DeJong, G. (1993). Assessing the Value of Information to Guide Learning Systems. In *Proceedings of the ML-93 Workshop on Knowledge Compilation and Speedup Learning*, pages 65–71, Tenth International Machine Learning Conference, University of Massachusetts, Amherst, MA.

Gratch, J., DeJong, G., & Chien, S.A. (1994). Deciding When and How to Learn. In M. desJardins & A. Ram, editors, *Proceedings of the AAAI Spring Symposium on Goal-Driven Learning*, pages 36–45, Stanford, CA.

Greeno, J.G. & Simon, H.A. (1988). Problem Solving and Reasoning. In R.C. Atkinson, H. Hernstein, G. Lindzey, & R.D. Luce, editors, *Stevens' Handbook of Experimental Psychology, Vol. 2: Learning and Cognition*, pages 589–673, Wiley, New York.

Hadzikadic, M. & Yun, D.Y.Y. (1988). Concept Formation by Goal-Driven, Context-Dependent Classification. In Z.W. Ras & L. Saitta, editors, *Methodologies for Intelligent Systems*, 3:322–332, 1988.

Hammond, K.J. (1989). *Case-Based Planning: Viewing Planning as a Memory Task*. Academic Press, Boston, MA.

Hammond, K., Converse, T., Marks, M., & Seifert, C.M. (1993). Opportunism and Learning. *Machine Learning*, 10(3):279–309.

Hayes-Roth, B. & Hayes-Roth, F. (1979). A Cognitive Model of Planning. *Cognitive Science*, 3(4):275-310.

Hayes-Roth, F. & Lesser, V. (1976). Focus of attention in a distributed logic speech understanding system. In *Proceedings of the IEEE International Conference on Accoustics, Speech and Signal Processing*, pages 416–420, Philadelphia, PA.

Hinrichs, T. (1992). *Problem Solving in Open Worlds: A Case Study in Design.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Hoffman, C., Mischel, W., & Mazze, K. (1981). The Role of Purpose in the Organization of Information about Behavior: Trait-based versus Goal-Based Categories in Human Cognition. *Journal of Personality and Social Psychology*, 39:211–255.

Hunter, L.E. (1990/chapter 2 of this volume). Planning to Learn. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, pages 261–276, Boston, MA.

Karlsson, J. (1994). Task Decomposition in Reinforcement Learning. In M. desJardins & A. Ram, editors, *Proceedings of the AAAI Spring Symposium on Goal-Driven Learning*, pages 46–53, Stanford, CA.

Kass, A. & Leake, D.B. (1988). Case-Based Reasoning Applied to Constructing Explanations. In J.L. Kolodner, editor, *Proceedings of the Case-Based Reasoning Workshop*, pages 190–208, Morgan Kaufmann Publishers, San Mateo, CA.

Keller, R. (1987). The Role of Explicit Contextual Knowledge in Learning Concepts to Improve Performance. Ph.D. Dissertation, Technical Report ML-TR-7, Department of Computer Science, Stanford University.

Keller, R. (1988). Defining Operationality for Explanation-Based Learning. *Artificial Intelligence*, 35:227–241.

Kedar-Cabelli, S. (1987). Formulating Concepts According to Purpose. In *Proceedings of the Sixth Annual National Conference on Artificial Intelligence*, pages 477-481, Seattle, WA.

Kintsch, W. (1988). The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. *Psychological Review*, 95(2):163–182.

Kocabas, S. (1994). Goal-Directed Discovery and Explanation in Particle Physics. In M. desJardins & A. Ram, editors, *Proceedings of the AAAI Spring Symposium on Goal-Driven Learning*, pages 54–61, Stanford, CA.

Kolodner, J.L. (1987). Extending Problem-Solving Capabilities through Case-Based Inference. In *Proceedings of the Fourth International Workshop on Machine Learning*, Irvine, CA.

Kolodner, J.L. (1993). *Case-Based Reasoning.* Morgan Kaufmann Publishers, San Mateo, CA.

Krulwich, B. (1994). A Component-Model Approach to Determining What to Learn. In M. desJardins & A. Ram, editors, *Proceedings of the AAAI Spring Symposium on Goal-Driven Learning*, pages 62–71, Stanford, CA.

Krulwich, B., Birnbaum, L., & Collins, G. (1992). Learning Several Lessons from one Experience. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, pages 242–247, Bloomington, IN.

Laird, J.E. (1991), editor. Special Section on Integrated Cognitive Architectures. *SIGART Bulletin*, 2(4):12–184.

Laird, J.E., Rosenbloom, P.S., & Newell, A. (1986). Chunking in Soar: The Anatomy of a General Learning Mechanism. *Machine Learning*, 1:11–46.

Leake, D.B. (1989). Anomaly Detection Strategies for Schema-Based Story Understanding. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society,* pages 490–497, Ann Arbor, MI.

Leake, D.B. (1991a). An Indexing Vocabulary for Case-Based Explanation. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, Anaheim, CA, pages 10–15.

Leake, D.B. (1991b/chapter 9 of this volume). Goal-Based Explanation Evaluation. *Cognitive Science*, 15:509–545.

Leake, D.B. (1992). *Evaluating Explanations: A Content Theory.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Leake, D.B. (1993). Learning Adaptation Strategies by Introspective Reasoning about Memory Search. In D.B. Leake, editor, *Proceedings of the AAAI-93 Workshop on Case-Based Reasoning*, pages 57–63, AAAI Press, Menlo Park, CA.

Leake, D.B. & Ram, A. (1993/chapter 16 of this volume). *Goal-Driven Learning: Fundamental Issues (A Symposium Report). AI Magazine*, 14(4).

Leake, D.B. (chapter 20 of this volume). Goal-Driven Integration of Explanation and Action. In A. Ram & D.B. Leake, editors, *Goal-Driven Learning*, MIT Press/Bradford Books, Cambridge, MA.

Maes, P. (1990). Situated Agents Can Have Goals. *Robotics and Autonomous Systems*, 6:49–70.

Markovitch, S. & Scott, P.D. (1993). Information Filtering: Selection Mechanisms in Learning Systems. *Machine Learning*, 10:113–151.

Martin, J.D. (1994). DC1: Supervised and Unsupervised Clustering. In M. desJardins & A. Ram, editors, *Proceedings of the AAAI Spring Symposium on Goal-Driven Learning*, pages 80–87, Stanford, CA.

Michalski, R.S. (1983). A Theory and Methodology of Inductive Learning. *Artificial Intelligence*, 20:111-161.

Michalski, R.S. (1993). Inferential Theory of Learning as a Conceptual Basis for Multistrategy Learning. *Machine Learning*, 11(2/3):111–151, 1993.

Michalski, R.S. & Ram, A. (chapter 21 of this volume). Learning as Goal-Driven Inference. In A. Ram & D.B. Leake, editors, *Goal-Driven Learning*, MIT Press/Bradford Books, Cambridge, MA.

Michalski, R.S. & Tecuci, G. (1994), editors. *Machine Learning: A Multistrategy Approach, Volume IV*, Morgan Kaufman Publishers, San Mateo, CA.

Minsky, M. (1963). Steps towards Artificial Intelligence. In E.A. Feigenbaum & J. Feldman, editors, *Computers and Thought*, pages 406–450, McGraw-Hill, New York, NY.

Minton, S. (1988). *Learning Search Control Knowledge: An Explanation-Based Approach.* Kluwer Academic Publishers, Boston.

Minton, S. (1990/chapter 3 of this volume). Quantitative Results Concerning the Utility of Explanation-Based Learning. *Artificial Intelligence*, 42:363-391.

Mitchell, T.M. (1980). *The Need for Biases in Learning Generalizations.* Technical Report CBM-TR-117, Rutgers University, New Brunswick, NJ.

Mitchell, T.M. (1982). Generalization as Search. *Artificial Intelligence*, 18:203–226.

Mitchell, T.M., Allen, J., Chalasani, P., Cheng, J., Etzioni, O., Ringuette, M. & Schlimmer, J. (1991). Theo: A Framework for Self-Improving Systems. In K. VanLehn, editor, *Architectures for Intelligence: The Twenty-Second Carnegie Mellon Symposium on Cognition*, pages 323–355, Lawrence Erlbaum Associates, Hillsdale, NJ.

Mitchell, T.M. & Keller, R. (1983). Goal Directed Learning. In *Proceedings of the International Machine Learning Workshop*, pages 117–118, Monticello, IL.

Mitchell, T.M., Keller, R., & Kedar-Cabelli, S. (1986). Explanation-Based Generalization: A Unifying View. *Machine Learning*, 1(1):47–80.

McKoon, G. & Ratcliff, R. (1992). Inference During Reading. *Psychological Review*, 99: 440-466.

Mooney, R. (1990). *A General Explanation-based Learning Mechanism and its Application to Narrative Understanding.* Morgan Kaufmann Publishers, San Mateo, CA.

Mooney, R.J. & Ourston, D. (1993). A Multistrategy Approach to Theory Refinement. In R.S. Michalski & G. Tecuci, editors, *Machine Learning: A Multistrategy Approach, Volume IV*, pages 141–164, Morgan Kaufman Publishers, San Mateo, CA.

Mostow, J. (1983). Machine Transformation of Advice into a Heuristic Search Procedure. In R.S. Michalski, J.G. Carbonell, and T.M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*, pages 367–403, Morgan Kaufmann Publishers, San Mateo, CA.

Mostow, J. & Bhatnagar, N. (1987). FAILSAFE—A Floor Planner that Uses EBG to Learn from its Failures. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pages 249–255, Milan, Italy.

Murphy, G.L. & Medin, D.L. (1985). The Role of Theories in Conceptual Coherence. *Psychological Review*, 92:289–316.

Newell, A. (1990). *Unified Theories of Cognition.* Harvard University Press. Cambridge, MA.

Newell, A. & Simon, H. (1972). *Human Problem Solving.* Prentice-Hall, Englewood Cliffs, NJ.

Ng, E. & Bereiter, C. (1991/chapter 14 of this volume). Three Levels of Goal Orientation in Learning. *The Journal of the Learning Sciences*, 1(3&4):243–271.

Oehlmann, R., Sleeman, D., & Edwards, P. (1992.) Self-Questioning and Experimentation in an Exploratory Discovery System. In *Proceedings of the ML-92 Workshop on Machine Discovery*, pages 41–50, Ninth International Machine Learning Conference, University of Aberdeen, Scotland.

Owens, C. (1991). A Functional Taxonomy of Abstract Plan Failures. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, pages 167–172, Chicago, IL.

Patalano, A., Seifert, C., & Hammond, K. (1993). Predictive Encoding: Planning for Opportunities. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, pages 800–805, Boulder, CO.

Park, Y. & Wilkins, D. (1990). Establishing the Coherence of an Explanation to Improve Refinement of an Incomplete Knowledge Base. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 511-516, Boston, MA.

Perez, M.A. (1994). The Goal is to Produce Better Plans. In M. desJardins & A. Ram, editors, *Proceedings of the AAAI Spring Symposium on Goal-Driven Learning*, pages 88–93, Stanford, CA.

Plaza, E., Aamodt, A., Ram, A., Van de Welde, W., & van Someren, M. (1993). Integrated Learning Architectures. In *Proceedings of the European Conference on Machine Learning*, Vienna, Austria.

Provost, F.J. (1994). Goal-Directed Inductive Learning: Trading off Accuracy for Reduced Error Cost. In M. desJardins & A. Ram, editors, *Proceedings of the AAAI Spring Symposium on Goal-Driven Learning*, pages 94–101, Stanford, CA.

Provost, F.J. & Buchanan, B.G. (1992). Inductive Policy. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 255–261, San Jose, CA.

Pryor, L. & Collins, G. (1992/chapter 10 of this volume). Planning to Perceive. In *Proceedings of the 1992 AAAI Spring Symposium on Selective Perception*, Stanford, CA.

Quilici, A. (in press). Toward Automatic Acquisition of an Advisory System's Knowledge Base. *Applied Intelligence.*

Rajamoney, S. (1993). Designing Experiments to Extend the Domain Theory. In DeJong, G., editor, *Investigating Explanation-Based Learning*, Kluwer, Boston.

Ram, A. (1987). AQUA: Asking Questions and Understanding Answers. In *Proceedings of the Sixth Annual National Conference on Artificial Intelligence*, pages 312–316, Seattle, WA.

Ram, A. (1989). *Question-Driven Understanding: An Integrated Theory of Story Understanding, Memory and Learning.* Ph.D. Dissertation, Research Report

#710, Yale University, Department of Computer Science, New Haven, CT.

Ram, A. (1990). Knowledge Goals: A Theory of Interestingness. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, pages 206–214, Cambridge, MA.

Ram, A. (1991). A Theory of Questions and Question Asking. *The Journal of the Learning Sciences*, 1(3&4):273–318.

Ram, A. (1993). Indexing, Elaboration and Refinement: Incremental Learning of Explanatory Cases. *Machine Learning*, 10:201–248.

Ram, A. & Cox, M.T. (1994/chapter 7 of this volume). Introspective Reasoning using Meta-Explanations for Multistrategy Learning. In R.S. Michalski & G. Tecuci, editors, *Machine Learning: A Multistrategy Approach, Volume IV*, Morgan Kaufman Publishers, San Mateo, CA.

Ram, A., Cox, M.T., & Narayanan, S. (chapter 18 of this volume). Goal-Driven Learning in Multistrategy Reasoning and Learning Systems. In A. Ram & D.B. Leake, editors, *Goal-Driven Learning*, MIT Press/Bradford Books, Cambridge, MA.

Ram, A. & Hunter, L. (1992/chapter 4 of this volume). The Use of Explicit Goals for Knowledge to Guide Inference and Learning. *Applied Intelligence*, 2(1):47–73.

Ram, A. & Leake, D.B. (1991). Evaluation of Explanatory Hypotheses. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, pages 867–871, Chicago, IL.

Ram, A., Narayanan, S., & Cox, M.T. (1993). Learning to troubleshoot: Multistrategy learning of diagnostic knowledge for a real-world problem solving task. Technical Report GIT-CC-93/67, College of Computing, Georgia Institute of Technology, Atlanta, Georgia 30332-0280, 1993.

Redmond, M. (1992). *Learning by Observing and Understanding Expert Problem Solving*. Ph.D. Dissertation, Technical report GIT-CC-92/43, College of Computing, Georgia Institute of Technology, Atlanta, GA.

Resnick, L.B. (1983). Towards a Cognitive Theory of Instruction. In S.G. Paris, G.M. Olson, & H.W. Stevenson, editors, *Learning and Motivating in the Classroom*, pages 6–38, Lawrence Erlbaum Associates, Hillsdale, NJ.

Resnick, L.B. (1987). Constructing Knowledge in School. In L.S. Liben, editor, *Development and Learning: Conflict or Congruence*, pages 19-50, Lawrence Erlbaum Associates, Hillsdale, NJ.

Riesbeck, C. (1981). Failure-driven Reminding for Incremental Learning. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pages 115–120, Vancouver, British Columbia.

Roschelle, J. (1992). Learning by Collaborating: Convergent Conceptual Change. *The Journal of the Learning Sciences*, 2(3):235–276.

Rosenbloom, P., Laird, J., & Newell, A. (1993) *The Soar papers: Research on integrated intelligence*. MIT Press, Cambridge, MA.

Sacerdoti, E.D. (1975). *A Structure for Plans and Behavior*. Technical Note #109, SRI International. Summarized in P.R. Cohen & E.A. Feigenbaum, *Handbook of AI, Volume III*, pages 541–550.

Scardamalia, M. & Bereiter, C. (1991). Higher Levels of Agency for Children in Knowledge Building: A Challenge for the Design of New Knowledge Media. *The Journal of the Learning Sciences*, 1(1):37–86.

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

Schank, R.C. (1986). *Explanation Patterns: Understanding Mechanically and Creatively*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Schank, R.C. & Abelson, R. (1977). *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Schank, R.C., Fano, A., Bell, B., & Jona, K. (1993/1994). The Design of Goal-Based Scenarios. *The Journal of the Learning Sciences*.

Schank, R.C. & Leake, D.B. (1989). Creativity and Learning in a Case-Based Explainer. *Artificial Intelligence*, 40(1–3):353–385. Also appears in J.G. Carbonell, editor, *Machine Learning: Paradigms and Methods*, MIT Press, Cambridge, MA, 1990.

Schlimmer, J.C. (1987). Incremental Adjustment of Representations for Learning. In *Proceedings of the Fourth International Workshop on Machine Learning*, pages 79–90, Irvine, CA.

Seifert, C. (1988). Goals in Reminding. In J.L. Kolodner, editor, *Proceedings of the Case-Based Reasoning Workshop*, pages 190–208, Morgan Kaufmann Publishers, San Mateo, CA.

Slade, S. (1993). *Goal-Based Decision Making: An Interpersonal Model*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Spellman, B.A. & Holyoak, K.J. (1993). An Inhibitory Mechanism for Goal-Directed Analogical Mapping. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, pages 947–952, Boulder, CO.

Sperber, D. & Wilson, D. (1986). *Relevance: Communication and Cognition*. Language and Thought Series, Harvard University Press, Cambridge, MA.

Srull, T. & Wyer, R. (1986). The Role of Chronic and Temporary Goals in Social Information Processing. In R. Sorrentino and E. Higgins, editors, *Handbook of Motivation and Cognition: Foundations of Social Behavior*, pages 503–549. Guilford Press, Guilford, CT.

Steinbart, P.J. (1992). The Role of Questioning in Learning from Computer-Based Decision Aids. In T.W. Lauer, E. Peacock, and A.C. Graesser, editors, *Questions and Information Systems*, pages 273–285, Lawrence Erlbaum Associates, Hillsdale, NJ.

Stroulia, E. & Goel, A.K. (1994). Task Structures: A Language for Learning Goals. In M. desJardins & A. Ram, editors, *Proceedings of the AAAI Spring Symposium on Goal-Driven Learning*, pages 112–121, Stanford, CA.

Stroulia, E., Shankar, M., Goel, A.K., & Penberthy, L. (1992). A Model-Based Approach to Blame Assignment in Design. In J.S. Gero, editor, *Proceedings of the Second International Conference on AI in Design*, pages 519–537.

Sussman, G. (1975). *A Computer Model of Skill Acquisition.* Artificial Intelligence Series, Volume 1, American Elsevier, New York, NY.

Tambe, M., Newell, A., & Rosenbloom, P.S. (1990). The Problem of Expensive Chunks and its Solution by Restricting Expressiveness. *Machine Learning*, 5:299-348.

Thagard, P. & Millgram, E. (chapter 19 of this volume). Inference to the Best Plan: A Coherence Theory of Decision. In A. Ram & D.B. Leake, editors, *Goal-Driven Learning*, MIT Press/Bradford Books, Cambridge, MA.

Utgoff, P.E. (1986). Shift of Bias for Inductive Concept Learning. In R.S. Michalski, J.G. Carbonell, & T.M. Mitchell, editors, *Machine Learning II: An Artificial Intelligence Approach*, Morgan Kaufman Publishers, San Mateo, CA.

van Berkum, J.J.A., Hijne, H., de Jong, T., van Joolingen, W.R., & Njoo, M. (1991/chapter 15). Aspects of Computer Simulations in an Instructional Context. *Education and Computing*, 6:231–239, 1991.

Van de Velde, W. (1988). Quality of Learning. In *Proceedings of the European Conference on Artificial Intelligence*, pages 408–413, Munich, Germany.

VanLehn, K. (1989). Problem Solving and Cognitive Skill Acquisition. In M.I. Posner, editor, *Foundations of Cognitive Science*, pages 527–579, MIT Press, Cambridge, MA.

VanLehn, K. (1991a). Rule Acquisition Events in the Discovery of Problem Solving Strategies. *Cognitive Science*, 15(1):1–47.

VanLehn, K. (1991b). *Architectures for Intelligence: The Twenty-Second Carnegie-Mellon Symposium on Cognition.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Veloso, M. & Carbonell, J.G. (1993). Derivational Analogy in PRODIGY: Automating Case Acquisition, Storage, and Utilization. *Machine Learning*, 10(3):249–278.

Weinert, F.E. (1987). Introduction and Overview: Metacognition and Motivation as Determinants of Effective Learning and Understanding. In F.E. Weinert & R.H. Kluwe, editors, *Metacognition, Motivation, and Understanding*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Weintraub, M. (1991). *An Explanation-Based Approach to Assigning Credit,* Ph.D. Dissertation, Computer Science Department, The Ohio State University, Columbus, OH.

Wellman, H.M. (1985). The Origins of Metacognition. In D.L. Forrest-Pressley, G.E. MacKinnon, & T.G. Waller, editors, *Metacognition, Cognition, and Human Performance*, Volume 1, pages 1–31, Academic Press, New York.

Wellman, H.M. (1992). *The Child's Theory of Mind.* MIT Press/Bradford Books, Cambridge, MA.

White, R.T. & Gunstone, R.F. (1989). Metalearning and Conceptual Change. *International Journal of Science Education*, 11:577-586.

Wilensky, R. (1983). *Planning and Understanding.* Addison-Wesley, Reading, MA.

Wisniewski, E.J. & Medin, D.L. (1991/chapter 6 of this volume). Harpoons and Long Sticks: The Interaction of Theory and Similarity in Rule Induction. In D. Fisher & M.J. Pazzani, editors, *Concept Formation: Knowledge and Experience in Unsupervised Learning*, Morgan Kaufman Publishers, San Mateo, CA.

Wrobel, S. (1988). Automatic Representation Adjustment in an Observational Discovery System. In D. Sleeman, editor, *Proceedings of the Third European Working Session on Learning*, pages 253–261, Pitman Publishing, London.

Xia, X. & Yeung, D.-Y. (1988/chapter 12). A Learning Model for the Selection of Problem Solving Strategies in Continuous Physical Systems. In *Proceedings of the Fourth IEEE Conference on Artificial Intelligence Applications*, pages 200–206, San Diego, CA.

Zukier, H. (1986). The Paradigmatic and Narrative Modes in Goal-Guided Inference. In R. Sorrentino and E.T. Higgins, editors, *Handbook of Motivation and Cognition: Foundations of Social Behavior*, Guilford Press, New York, NY.