Invited presentation at SOFSEM'96; appeared in LNCS 1175

Learning in Order to Reason: The Approach

Dan Roth*
Dept. of Appl. Math. & CS,
Weizmann Institute of Science
Rehovot 76100
Israel
danr@wisdom.weizmann.ac.il
http://www.wisdom.weizmann.ac.il/~danr

Abstract

Any theory aimed at understanding *commonsense* reasoning, the process that humans use to cope with the mundane but complex aspects of the world in evaluating everyday situations, should account for its flexibility, its adaptability, and the speed with which it is performed. Current theories of reasoning, however, do not satisfy these requirements, a fact we attribute, at least partly, to their separation from learning.

While the central role of learning in cognition is widely acknowledged, most lines of research nevertheless study the phenomenon of "learning" separately from that of "reasoning". The work presented here is motivated by the belief that learning is at the core of any attempt at understanding high level cognitive tasks. A formal model for the study of reasoning is developed in which a learning component has a principal role, and its advantages over traditional formalisms for the study of reasoning are shown.

This paper presents an integrated theory of learning, knowledge representation and reasoning within a unified framework called *Learning to Reason.* The Learning to Reason framework combines the interfaces to the world used by known learning models with a reasoning task and a performance criterion suitable for it. It is shown that the framework efficiently supports "more reasoning" than traditional approaches and at the same time matches our expectations of plausible patterns of reasoning. Several results are presented to substantiate this claim, presenting cases where learning to reason about the world is feasible but either reasoning from a given representation of the world or learning representations of the world do not have efficient solutions.

Overall, this framework suggests an "operational" approach to reasoning, that is nevertheless rigorous and amenable to analysis. As such, it may be a step toward a rigorous large-scale empirical study of learning and reasoning.

The paper presents work originally introduced by Khardon and Roth (Khardon & Roth 1994a) and surveys further developments made within this framework more recently.

*Supported by a grant from the Israeli Ministry of Science

Introduction

Consider a baby robot, starting out its life. If it were a human being, nature would have provided for the infant a safe environment in which it can spend an initial period of time. In this period the robot adapts to its environment and learns about the structures, rules, meta-rules, superstitions and other information the environment provides. In the meantime, the environment protects it from fatal events. Only after this "grace period", is the robot expected to have "full functionality" in the environment, but naturally, its performance depends on this environment and reflects the amount of interaction it has had with it.

While the central role of learning in cognition is widely acknowledged, early theories of intelligent systems have assumed that cognition (namely, computational processes such as reasoning, language understanding, object recognition and other "high level" cognitive tasks) can be studied separately from learning, or as phrased by Kirsh (Kirsh 1991), that "learning can be added later".

This paper presents a new framework for the study of Reasoning. In contrast to earlier approaches to reasoning, the *Learning to Reason* framework views learning as an integral part of the process, and suggests to study the entire process of *learning* some knowledge representation and *reasoning* with it.

In this framework an agent is given access to its favorite learning interface, and is also given a grace period in which it can interact with this interface and construct a representation KB of the world W. The reasoning performance is measured only after this period, when the agent is presented with its reasoning task. A related scenario in which the agent learns and reasons in an on-line fashion is also studied and sometimes yields a more natural view of the learning and reasoning process.

In the Learning to Reason framework it is not assumed that the knowledge representation describing the "world" is *given* to the agent. Instead, the agent constructs the knowledge representation while interacting with the world. In this way the reasoning task is no longer a "stand alone" process, and the agent does not need to reason from a previously defined "general purpose" knowledge representation. Rather, it can choose a knowledge representation that facilitates the reasoning task at hand. Moreover, we take the view that a reasoner need not answer efficiently *all* possible queries, but only those that are "relevant", or "common", in a well defined sense. This relaxation can be used by the agent in selecting its knowledge representation. In addition, by viewing the interaction of the agent with the environment while learning and reasoning in a unified way the performance of the agent can be measured relative to the environment it interacts with. Thus, while in the Learning to Reason framework the knowledge representation used by the agent is still a crucial ingredient, its effectiveness now depends on whether it is efficiently learnable and, at the same time, supports efficient reasoning performance relative to the environment.

We prove the usefulness of the Learning to Reason approach by showing that through interaction with the world, the agent truly gains additional reasoning power, over what is possible in the traditional setting. Several results are presented to substantiate this claim, exhibiting cases where learning to reason about the world is feasible but either (1) reasoning from a given representation of the world or (2) learning representations of the world do not have efficient solutions.

In this paper we present a high level survey of the theoretical work within the Learning to Reason framework. The work on this framework started by Khardon and Roth in (Khardon & Roth 1994a), and many of the works discussed here are extensions of this paper in various directions. No technical details are given here. Rather, we motivate the framework, describe its high level principles and briefly discuss how they can be implemented and what results they yield. For preliminaries on reasoning, learning and rigorous definitions for the material presented here, consult (Khardon & Roth 1994a).

Motivation

The generally accepted framework for the study of reasoning in intelligent systems is the knowledge-based system approach (McCarthy 1958; Nilsson 1991). It is assumed that the knowledge is given to the system, stored in some *representation language* with a well defined meaning assigned to its sentences. The sentences are stored in a Knowledge Base (KB) which is combined with a reasoning mechanism, used to determine what can be inferred from the sentences in the KB. Many knowledge representations can be used to represent the knowledge in a knowledge-based system. Different representation systems (e.g., a set of logical rules, a probabilistic network) are associated with corresponding reasoning mechanisms, each with its own merits and range of applications. The question of how this knowledge might be acquired and whether this should influence how the performance of the reasoning system is measured is normally not considered. The intuition behind this approach is based on the following observation:

Observation: *If there is a learning procedure that can learn an exact description of the world in representation R, and there is a procedure that can reason exactly using R, then there is a complete system that can learn to produce "intelligent behavior" using R.*

We believe that the separate study of learning and the rest of cognition is, at least partly, motivated by the assumption that the converse of the above observation also holds. Namely, that if there is a system that can Learn to Reason, then there is a learning procedure that can learn a representation of the world, and a reasoning procedure that can reason with it.

Computational considerations, however, render the traditional self-contained reasoning approach as well as other variants of it inadequate for common-sense reasoning. This is true not only for the task of deduction, but also for many other forms of reasoning which have been developed, partly in order to avoid the computational difficulties in exact deduction and partly to meet some (psychological and other) plausibility requirements. All those were shown to be even harder to compute than the original formulation (Selman 1990; Papadimitriou 1991; Roth 1996). As a consequence, many recent works in reasoning aim at identifying classes of limited expressiveness, with which one can perform some sort of reasoning efficiently (Levesque & Brachman 1985; Cadoli 1995; Levesque 1992; Selman 1990). However, none of these works meet the strong tractability requirements for common-sense reasoning (as described, for example, in (Shastri 1993)), even though, (as argued, for example, in (Doyle & Patil 1991)) the inference is sometimes restricted in implausible ways.

Very few works have considered the question of integrating theories of reasoning and learning in any formal way. In fact, results in these two fields are in a fairly disconnected state. The current emphasis of the research in learning is on the study of inductive learning (from examples) of concepts (binary classifications of examples). In this framework the performance of the learner is measured when classifying future examples. Perhaps the most important open question in learning theory today is concerned with the learnability of DNF or CNF formulas (the problems are equivalent in the current framework). However, even if one had a positive result for the learnability of these classes, this would be relevant only for classification tasks, and cannot be used for reasoning. The reason is that if the output of the learning algorithm is a CNF expression, then it cannot be used for reasoning, since this problem is computationally hard. From a traditional reasoning point of view, on the other hand, learning a DNF is not considered interesting, since it does not relate easily to a rule based representation. Alternative representations studied in learning theory are also not geared towards supporting the reasoning task, and are thus not directly usable. Other problems that exist in the interface between a learning algorithm and a reasoning algorithm are discussed later in this paper.

In this work, therefore, while we build on the framework and some of the results of computational learning theory, we distinguish the traditional learning task which we call here *Learning to Classify* (L2C) from the new learning task, *Learning to Reason*.

The Learning to Reason approach should also be contrasted with various knowledge compilation studies (Selman & Kautz 1991; Moses & Tennenholtz 1993). There, a theory (KB) is given to the system designer who is trying to compile it, off line, into a more tractable knowledge representation, to facilitate the answering of future queries. In our approach, a world representation is not given to the agent, but instead, it is assumed that the agent can access the world itself via some reasonable interface and acquire information that, later on, will support query answering correctly and efficiently.

This work is similar in nature to the Neuroidal model developed by Valiant (Valiant 1994). The model developed there provides a more comprehensive approach to cognition, and akin to our approach it views learning as an integral and crucial part of the process. There, the agent reasons from a learned knowledge base, a complex circuit, and thus can be modeled by our framework. Indeed reasoning in the Neuroidal model shares many properties with the Learning to Reason framework. One difference is that in some instances of the Learning to Reason framework, though not all, we restrict our discussion to a fixed, consistent world, in an effort to give a more formal treatment of a reasoner that has learned its knowledge base.

Learning to Reason

Motivated by the abovementioned computational considerations we argue that a central question to consider, if one wants to develop computational models for commonsense reasoning, is how the intelligent system acquires its knowledge and how this process of interaction with its environment influences the performance of the reasoning system. Thus the Learning to Reason theory is concerned with studying the entire process of *learning* some knowledge representation and *reasoning* with it. In its most abstract form the Learning to Reason approach has the following principles:

- Intelligent agents are not omniscient:
  The view of commonsense reasoning taken here is that the agent has to function in a very complex world that may be hard to represent exactly. Luckily, the agent need not be omniscient, but rather has to perform well on a fairly wide, but restricted, set of tasks. Thus, the requirements from the reasoning stage may be relaxed.

- The goal of the learning stage depends on the required functionality:
  The learning stage is not evaluated by how well its output models the world, but rather by how well it supports the required functionality. Given, for example, that the agent is only required to perform well on a restricted[1] class of tasks, there may not be a need for the agent to learn a complete description of the world. A partial or approximate representation may be sufficient to support the relaxed reasoning requirements.

- Interaction with the world is a key issue:
  The interaction of the agent with its environment during the learning stage is an important aspect of this view. The type of interaction assumed depends on the task the agent is to perform and may range from observing examples, actively studying the environment using membership queries, interacting with a teacher or "being told" some facts. Naturally, there may be a tradeoff between the strength of this interaction and the resulting functionality.

- The knowledge representation used may depend on the functionality:

[1]This should not be taken too narrowly. The intention is not to perform well a single mission, but rather learn in order to perform well on a fairly wide collection of tasks, which share some commonality. This is the way it has been used in the cases already studied.

The notion of knowledge representation is as important in the Learning to Reason framework as in the more traditional KBS framework. However, the effectiveness of the knowledge representation here depends on its learnability and on how well it supports inference, rather than on its comprehensibility. In this way, there may not exist a "general purpose" knowledge representation on which a "general purpose" inference engine can act. Instead, different knowledge representations should learned in order to support various tasks.

- The performance of the agent is measured with respect to the world it functions in, and not in any absolute terms:

  The world in which the agent performs its task is the same world that supplies the agent the information when learning. One interpretation of this principle may be that the performance of the agent is measured only on a collection of tasks that are "relevant" or "common" in the environment.Another may be that the same arbitrary "world" that supplies the information in the learning phase is used to measure the agent's performance later. In its general form this principle induces a unified way to view the interaction of the agent with its environment during the learning and reasoning stages, and suggests that both should be governed by the same distribution.

  In general, it may not be necessary to appeal to a notion of a "world" at all (e.g., by not making any assumptions on the world the agent functions in) when the performance of the agent on the required functionality can be measured with respect to the functionality observed while learning.

- Rigor and efficiency:

  The aim was to define the framework in a way that is rigorous and amenable to analysis. For this purpose the interaction of the agent with its world is defined in a formal way (as in Computational Learning Theory), as are the tasks to be performed and any assumptions made on the world the agent functions in. In addition, it is usually required that Learning to Reason is done in time that is polynomial in the natural complexity parameters.

Results

Reasoning, as the term is used in AI, is viewed as having a major role in several high level cognitive tasks, including language understanding, high level vision and planning, tasks which rely on performing some sort of *inference*. A basic inference task considered in this context, that of *deductive inference*, is the focus of this presentation. In the first part of this section we concentrate on the ideas as presented by Khardon and Roth in the original paper on this framework, and discuss some of the results proved there. Later, we briefly survey some other results proved within this framework. The results are presented at a high level and without any of the technical details. Consult the relevant papers for those.

General Framework

Several general questions regarding the relation of the Learning to Reason (L2R) framework to the two existing ones, the traditional reasoning framework and the traditional learning framework have been considered (Khardon & Roth 1994a). Most of these were considered for the task of deductive reasoning. First, it was shown that when the class of queries is not restricted, L2R implies L2C. However, the interesting aspect of this is that this property does not hold if the class of queries is restricted in some natural way (see next section). A second basic question to consider concerns the possibility of Learning to Reason by putting together existing learning and reasoning algorithms. As pointed out earlier, this approach has a problem whenever the output of the learning algorithm does not support efficient reasoning, as happens in many of the commonly used knowledge representations. Even when this is not a problem, it turns out that the straightforward approach that builds a L2R system by reasoning from the output of a (PAC or mistake bound) L2C algorithm, has some shortcomings. In particular, it is shown that a PAC learning algorithm, provided that it has an additional property ("learning from below"), can be combined with a reasoning algorithm to yield a PAC-Learn to Reason algorithm. The significance of this result, however, is that it exhibits the *limitations* of L2R by combining reasoning and learning algorithms: relaxing the requirement that the algorithm learns from below is not possible. Similar behavior is shown for mistake-bound algorithms.

Deductive Reasoning

The most striking evidence of the usefulness of this approach is given in the context of deductive reasoning. It is shown that the new framework allows for efficient solutions even in cases where the separate learning and reasoning tasks are not tractable. The following results are shown for Learning to Reason algorithms that use a set of models (satisfying assignments) as their knowledge representation. The results build on a characterization of reasoning with models developed in (Khardon & Roth 1994b) (based on ideas from (Bshouty 1995)).

- **Learning to Reason without Reasoning:**
Consider the reasoning problem W $\models \alpha$, where W is some CNF formula and α is a $\log n$CNF (i.e., a CNF formula with at most $\log n$ literals in each clause). Then, when W has a polynomial size DNF[2] there is an exact and efficient Learning to Reason algorithm for this problem, while the traditional reasoning problem (with a CNF representation as the input) is NP-Hard.

- **Learning to Reason without Learning to Classify:**
Consider the reasoning problem W $\models \alpha$, where W is any Boolean formula with a polynomial size DNF and α is a $\log n$CNF. Then, there is an exact and efficient Learning to Reason algorithm for this problem, while the class of Boolean formulas with polynomial size DNF is not known to be learnable in the traditional (Learning to Classify) sense.

Learning to Reason algorithms that use formulas as their knowledge representation are also considered, and results of the same nature can be shown there too. In the following result the formula-based knowledge representation does not describe the world exactly but rather an approximation of it (see below). The following builds on a learning (to classify) result of Frazier and Pitt (1993):

- **Learning to Reason without Reasoning:**
Consider the reasoning problem W $\models \alpha$, where W is any Boolean formula that has a Horn approximation of polynomial size, and α is a Horn expression. Then, there is an exact and efficient Learning to Reason algorithm for this problem, while the problem of learning W exactly is not known to be solvable, and the problem of reasoning from a representation of W is not tractable.

Of course, these algorithms do not solve NP-hard problems. Rather, the additional reasoning power is gained through the interaction with the world. In the first instance, examples from the world are used to construct the model-based representation. In the second instance the queries presented by the interface are used to construct the approximation of W. An additional crucial observation, used in two of these results is that in order to reason with respect to W one need not learn W exactly. Instead, it is sufficient to use the least upper bound approximation of W. (The least upper bound approximation is, in some sense, (Selman & Kautz 1996; Khardon & Roth 1994b) the function closest to W in the class of queries we reason about). These approximations are shown to be learnable in a form that supports the reasoning task efficiently, and this is used to prove the Learning to Reason results.

These results show that neither a traditional reasoning algorithm (from the CNF representation) nor a traditional learning algorithm (that can "classify" the world) is necessary for Learning to Reason. Moreover, the results exemplify the phrase "intelligence is in the eye of the beholder" (Brooks 1991), since our agent seems to behave logically, even though its knowledge representation need not be a logical formula and it does not use any logic or "theorem proving".

To summarize, the new positive results are made possible by a combination of several features, which can be viewed as a direct application of the general principles listed above to the current instantiation. First, we relax the inference problems by restricting the classes of queries considered[3], while, at the same time, using different knowledge representations (that may not be in the traditional comprehensible form) in which this can be exploited. Second, we represent in our KB the least upper bounds of the "world" function rather than the exact representation. Third, and perhaps conceptually most important, our formal framework for the study of reasoning is different from previous ones since we allow the agent to interact with the world, and can therefore measure its performance relative to the world.

Other Learning to Reason Results

As mentioned above, the framework should be seen in a more general context and can be applied in a variety of tasks. We briefly point to results which have be recently developed for other, related, reasoning tasks within this framework. We discuss in the following only theoretical results within this framework, and do not consider more applied work that is influenced by this framework (Golding & Roth 1996).

Abductive Reasoning

The results cited above are based on learnability results for model-based representations. Together with the results in (Khardon & Roth 1994b), which show how model-based representations can be used for efficient abductive reasoning (see there for details on the abduction formalisms used) this yields an algorithm for Learning to Reason abductively. Moreover, as in

[2]The DNF representation is not given to the reasoner. Its existence is essential, since the algorithm is polynomial in its size.

[3]Notice that restricting the classes of queries considered does not change the intractability of the deduction problem, if the world is represented traditionally, as a CNF formula.

the deductive case, the result obtained can be phrased as a "Learning to Reason without Reasoning" result.

Default Reasoning

As in the case of abductive reasoning, learnability results for model-based representations, together with the results in (Khardon & Roth 1995a), which show how model-based representations can be used for efficient default reasoning, yield an algorithm for Learning to Reason with defaults. In particular, the results provide a "Learning to Reason without Reasoning" result to fragments of Reiter's default logic.

Reasoning with Partial Assignments

The deductive reasoning approach presented above has been extended in (Khardon & Roth 1995b) to handle partial assignments in the input. Several interpretations for partial information in the interface with the environment are discussed there and the work on model-based representations is extended to deal with partially observable worlds. Then, learning to reason algorithms that cope with partial information are presented. These results exhibit a tradeoff between learnability, the strength of the oracles used in the interface and the expressiveness of the queries asked. As in the cases above, it is shown that one can learn to reason with respect to expressive worlds, that cannot be learned efficiently in the traditional learning framework, and do not support efficient reasoning in the traditional reasoning framework.

In addition, this work suggests another important motivation for the study of reasoning (and in particular, deductive reasoning) and for integrating it with learning. It is shown that when dealing with partial information in the interface, classification problems become deductive reasoning problems.

Non-Monotonic Reasoning

In (Roth 1995; Valiant 1995) a different view of reasoning in the presence of partial assignments is developed. The approach presented there implements in its general form the L2R principle that the performance of an agent is measured with respect to the world it functions in. Namely, the interaction of an agent with its environment during the learning and reasoning stages are defined in a unified way, via the notion of an *observation.*

This is used to formalize the intuition that incomplete information may actually help to support efficient and plausible reasoning; the underlying assumption is that missing information in the interaction of the agent with its environment may be as informative for future interactions as observed information.

Formally, (Roth 1995) shows that the problem of reasoning from incomplete information can be presented as a problem of learning attribute functions over a generalized domain. Several examples, which have been used over the years as benchmarks for various formalisms and that illustrate various aspects of the non-monotonic reasoning phenomena, are considered and translated into Learning to Reason problems. It is then demonstrated that these have concise representations over the generalized domain and it is shown that these representations can be learned efficiently, yielding Learning to Reason algorithm that learn to reason non-monotonically.

Learning to Take Action

(Khardon 1996) extends the framework in another direction and studies planning problems. As in other instances of the Learning to Reason framework, the problem of learning to take actions is viewed as a supervised learning problem. In this case, the learning problem is in a dynamic stochastic domain; the agent receives observations (a teacher acting in the world) and learns from it an acting strategy. This model implements the L2R principles and, in particular, the performance of the agent is measured with respect to the world it functions in with very few assumptions made on the world. The knowledge representation selected in this case is that of production rule systems and it is shown that action strategies based on this representation can be learned. The most significant addition to the framework developed there is that the agent *acts* in the world, and there by changes it. Other works in planning which can be viewed within the Learning to Reason framework include, for example, (Baum 1996).

Learning Active Classifiers

Many classification algorithms are "passive", in that they assign a class-label to each instance based only on the description given, even if that description is incomplete. In contrast, an *active* classifier can, at some cost, obtain the values of missing attributes, before deciding upon a class label. The problem of learning active classifiers is formalized and studied in (Greiner, Grove, & Roth 1996). It is shown there that while the "learn then optimize" approach to this problem is certainly sufficient (in principle) to determine active classifiers, it can fail (for complexity reasons) in various ways. Perhaps the main point made there is that one may be better off learning the active classifier directly. The basic idea implements some of the L2R principles, in that it is suggested to learn just enough to perform some particular task, in a representation tailored to this task, rather than trying to learn everything.

Conclusions

We have presented the *Learning to Reason* framework, a recently introduced framework for the study of reasoning in intelligent systems, and surveyed some of the recent results shown within it.

The *Learning to Reason* approach is intended to overcome some of the fundamental problems in earlier approaches to reasoning. This framework differs from existing ones in that it sees learning as an integral part of the process, it avoids enforcing rigid syntactic restrictions on the intermediate knowledge representation, and it makes explicit the dependence of the reasoning performance on the input from the environment.

The usefulness of the Learning to Reason approach is shown by exhibiting a few interesting results, that are not possible in the traditional setting. For the problem of deductive reasoning we have shown cases in which the new framework allows for successful Learning to Reason algorithms, but stated separately, either the reasoning problem or the learning problem are not (or not known to be) tractable. Results of the same nature have been shown also for other inference problems including various reasoning and planning formalisms.

We have made explicit the main principles of the approach and have demonstrated that these can be implemented in various ways in many inference problems. In all these cases we have shown that the Learning to Reason approach efficiently supports "more reasoning" than traditional approaches and at the same time matches our expectations of plausible patterns of reasoning.

Certainly, these are not the only works which can be viewed as implementations of the L2R principles. Some practitioners have argued before, as is argued here, for an "operational" approach to the study of reasoning. One of the contributions of this line of research is that it shows, in a formal sense, that an operational approach is not a "necessary evil" but rather a well justified path and moreover, that an "operational" approach to reasoning can be developed, that is rigorous and amenable to analysis.

We believe that this framework is a step toward constructing an adequate computational theory of reasoning. One major difference between the traditional, knowledge-based system approach to intelligent inference and the L2R approach is that the latter approach suggests that for large scale reasoning to work in practice, reasoning systems need to be trained over a large number of examples. Integrating the knowledge acquisition stage with the reasoning stage in a plausible manner, as suggested here, may thus be an important step toward a rigorous large-scale empirical study of learning and reasoning.

References

Baum, E. 1996. Toward a model of mind as a laissez-faire economy of idiots. In *Proceedings of the International Conference on Machine Learning.*

Brooks, R. A. 1991. Intelligence without representation. *Artificial Intelligence* 47:139–159.

Bshouty, N. H. 1995. Exact learning via the monotone theory. *Information and Computation* 123(1):146–153.

Cadoli, M. 1995. *Tractable Reasoning in Artificial Intelligence.* Springer-verlag. Lecture notes in Artificial Intelligence, vol. 941.

Doyle, J., and Patil, R. 1991. Two theses of knowledge representation: language restrictions, taxonomic classification, and the utility of representation services. *Artificial Intelligence* 48:261–297.

Frazier, M., and Pitt, L. 1993. Learning from entailment: An application to propositional Horn sentences. In *Proceedings of the International Conference on Machine Learning.* Morgan Kaufmann.

Golding, A. R., and Roth, D. 1996. Applying winnow to context-sensitive spelling correction. In *Proceedings of the International Conference on Machine Learning.*

Greiner, R.; Grove, A.; and Roth, D. 1996. Learning active classifiers. In *Proceedings of the International Conference on Machine Learning*, 207–215.

Khardon, R., and Roth, D. 1994a. Learning to reason. In *Proceedings of the National Conference on Artificial Intelligence*, 682–687.

Khardon, R., and Roth, D. 1994b. Reasoning with models. In *Proceedings of the National Conference on Artificial Intelligence*, 1148–1153. To appear in Artificial Intelligence Journal.

Khardon, R., and Roth, D. 1995a. Default-reasoning with models. In *Proceedings of the International Joint Conference of Artificial Intelligence*, 319–325. To appear in Artificial Intelligence.

Khardon, R., and Roth, D. 1995b. Learning to reason with a restricted view. In *Workshop on Computational Learning Theory*, 301–310.

Khardon, R. 1996. Learning to take actions. In *Proceedings of the National Conference on Artificial Intelligence*, 787–792.

Kirsh, D. 1991. Foundations of AI: the big issues. *Artificial Intelligence* 47:3–30.

Levesque, H., and Brachman, R. 1985. A fundamental tradeoff in knowledge representation and reasoning. In Brachman, R., and Levesque, H., eds., *Readings in Knowledge Representation.* Morgan Kaufman.

Levesque, H. 1992. Is reasoning too hard ? In *Proceeding of the 3rd NEC research Symposium.*

McCarthy, J. 1958. Programs with common sense. In Brachman, R., and Levesque, H., eds., *Readings in Knowledge Representation, 1985.* Morgan-Kaufmann.

Moses, Y., and Tennenholtz, M. 1993. Off-line reasoning for on-line efficiency. In *Proceedings of the International Joint Conference of Artificial Intelligence*, 490–495.

Nilsson, N. J. 1991. Logic and artificial intelligence. *Artificial Intelligence* 47:31–56.

Papadimitriou, C. H. 1991. On selecting a satisfying truth assignment. In *Proc. 32nd Ann. IEEE Symp. on Foundations of Computer Science*, 163–169.

Roth, D. 1995. Learning to reason: The non-monotonic case. In *Proceedings of the International Joint Conference of Artificial Intelligence*, 1178–1184.

Roth, D. 1996. On the hardness of approximate reasoning. *Artificial Intelligence* 82(1-2):273–302.

Selman, B., and Kautz, H. 1991. Knowledge compilation using Horn approximations. In *Proceedings of the National Conference on Artificial Intelligence*, 904–909.

Selman, B., and Kautz, H. 1996. Knowledge compilation and theory approximation. *Journal of the ACM* 43(2):193–224.

Selman, B. 1990. *Tractable Default Reasoning.* Ph.D. Dissertation, Department of Computer Science, University of Toronto.

Shastri, L. 1993. A computational model of tractable reasoning - taking inspiration from cognition. In *Proceedings of the International Joint Conference of Artificial Intelligence*, 202–207.

Valiant, L. G. 1994. *Circuits of the Mind.* Oxford University Press.

Valiant, L. G. 1995. Rationality. In *Workshop on Computational Learning Theory*, 3–14.