From *Proceedings of the Fourteenth National Conference on Artificial Intelligence* (1997). Providence, RI: AAAI Press.

Machine Learning for Intelligent Systems

Pat Langley*
Intelligent Systems Laboratory
Daimler-Benz Research & Technology Center
1510 Page Mill Road, Palo Alto, CA 94304
(Langley@cs.stanford.edu)

Abstract

Recent research in machine learning has focused on supervised induction for simple classification and reinforcement learning for simple reactive behaviors. In the process, the field has become disconnected from AI's original goal of creating complete intelligent agents. In this paper, I review recent work on machine learning for planning, language, vision, and other topics that runs counter to this trend and thus holds interest for the broader AI research community. I also suggest some steps to encourage further research along these lines.

Introduction

A central goal of artificial intelligence has long been to construct a complete intelligent agent that can perceive its environment, generate plans, execute those plans, and communicate with other agents. The pursuit of this dream naturally led many researchers to focus on the component tasks of perception, planning, control, and natural language, or on generic issues that cut across these tasks, such as representation and search. Over the years, the AI field has gradually fragmented into many distinct communities, each concerned with a different facet of intelligent behavior.

This separate-and-conquer strategy has some benefits, since the researchers in each area can concentrate their energies and make more rapid progress on their own problems. But it has also led to differences in notation and terminology that make it difficult to communicate across paradigm boundaries. More important, it has led each subcommunity to focus on problems that are remote from the original goal of building a complete intelligent agent. Although there have been some exceptions to this tendency, most AI research now bears little resemblance to the original vision for the field.

In this paper, we focus on one area – machine learning – that exemplifies this trend. Many have argued (e.g., Langley & Simon, 1981) that learning has a central role to play in intelligent behavior, and these are arguments are no less valid today than in AI's earlier days. Yet research within the machine learning community has come to show little concern with the broader issues of intelligent systems. We begin by characterizing the form that this problem has taken in machine learning, and then review some counterexamples to the trend, showing that not all researchers have forgotten AI's original aims. We close by outlining some problems that have yet to be addressed and suggest some actions that we can take to counter the fragmentation process, at least with respect to learning research.

*Also affiliated with Institute for the Study of Learning and Expertise, 2164 Staunton Court, Palo Alto, CA 94306.

Developments in Machine Learning

Over the past decade, machine learning has developed into a well-defined discipline with clear goals that revolve around improving performance with experience. The field has strong experimental, theoretical, and applied arms, each with its own criteria for success. Most important, machine learning has substantially broadened its vision to incorporate a wide range of methods, some having their origins in AI but others coming from pattern recognition and statistics.

This developmental trend is certainly encouraging. In 1986, the journal *Machine Learning*'s first year, most researchers within the community acknowledged only certain methods, mainly those that learned logical rules or decision trees, as their object of study. By 1996, this picture had changed drastically, with papers in the journal and at the annual conference also dealing with methods that learn weights in multilayer neural networks, that store training cases in memory and generalize only at retrieval time, and that induce probabilistic descriptions rather than logical ones. Any methods that show evidence of improving performance with experience are now considered fair game for the machine learning researcher.

This period has seen considerable advances in methods for supervised learning for use on classification and prediction tasks. For example, early induction methods tended to overfit noisy training data, but more recent techniques for decision-tree induction, neural networks, case-based learning, and probabilistic induction incorporate schemes, such as pruning and weight decay, to counter this effect. There has also been progress on extending supervised methods to handle data with missing values, to deal with domains having many classes, and to support relational descriptions (often called 'inductive logic programming'). Other advances include methods for improving predictive accuracy by taking domain knowledge into account during induction, by reducing dimensionality through selecting useful features, and by combining classifiers learned with different methods or from different training sets.

There has also been considerable activity on 'reinforcement learning', which involves learning control strategies from delayed rewards. The typical formulation assumes a physical agent that makes reactive decisions about what action to take at each step, initially at random but with increasing selectivity as it gains experience. The agent can receive reward from the environment on each step, but usually gets no information until it has taken many actions. Most (though not all) work on this topic uses some form of temporal-difference method (Sutton, 1988), which propagates reward backward through time to better estimate the expected reward for taking a given action in a given state. Early work simply stored these scores in a large state-action table, but more recent studies have combined this approach with induction techniques that generalize across states to speed the learning process.

Unfortunately, as the machine learning community has broadened on some dimensions, it has narrowed on others. The present emphasis on supervised learning for simple classification and reinforcement learning for simple reactive behaviors makes little contact with the wider goals of artificial intelligence. The distribution of current research in machine learning differs markedly from that 15 years ago, when there was considerable work on learning for the core AI areas of natural language, planning, and perception. Concern with such complex tasks has been replaced with emphases on increasing the accuracy of supervised learning methods and making the learning rate in reinforcement learning more practical.[1]

However, this does *not* mean there has been less work on learning for natural language, planning, and perception. There has actually been considerable growth in the amount of research devoted to these topics, but this work has not been done by people who call themselves machine learning researchers. Rather, these efforts are due to researchers who specialize in natural language, computer vision, speech understanding, and planning.[2]

Yet even within these problem-oriented communities, there has been growing reliance on simple learning schemes. For example, within natural language, common tasks include learning to determine (classify) the sense of a word from surrounding context and where to attach a relative clause. Similar reformulations have occurred within the computer vision community, again leading to a focus on simple methods for supervised learning. Thus, both within the machine learning community and elsewhere, we see increasing interest in learning issues but a growing divide between the tasks addressed and the traditional concerns of artificial intelligence.

Some Enlightened Counterexamples

Fortunately, not all researchers have fallen prey to this trend, and some continue to study learning within the context of the larger-scale tasks that seem necessary for intelligent agents. In this section we review some exemplary work along these lines, taking our examples mostly from the machine learning community but organizing them in terms of problem areas.

Natural Language Learning

One important counterexample to the above trend is the research program that Zelle and Mooney (1993, 1996) have carried out on natural language. Whereas most of the natural language community has embraced probabilistic and statistical approaches that require large corpora for learning, Mooney and Zelle have instead worked on learning condition-action rules using inductive logic programming. Moreover, their Chill system relies not on N-grams or probabilistic grammars for their performance element, which make little contact with mainstream AI, but rather uses a shift-reduce parser. Such a parser operates like a production system, on each step selecting a rule that alters a parse tree or modifies a buffer.

At the outset, Chill's rules are overly general, allowing it to parse many ungrammatical sentences. But the system uses the parse trees associated with each training sentence to distinguish between legal and illegal applications of each rule, which become positive and negative training cases. From these, Chill learns more specific variants of the original rules that, typically, produce legal parses but not illegal ones. This approach has much in common with early work on learn-

[1] Some reasons for these changes are clear: the UCI data repository has made it very easy to run experiments with supervised methods, the corporate interest in data mining has rewarded work on classification learning, and the mathematical elegance of temporal-difference methods has attracted many adherents.

[2] This trend has held less true in planning and problem solving, where learning work has been about equally divided between the planning and machine learning communities.

ing search heuristics for problem solving (Sleeman, Langley, & Mitchell, 1983), which used solution paths found through search to generate positive and negative training cases.

Zelle and Mooney have tested this approach on corpora (sentences and hand labeled parse trees) used by researchers in the statistical language community, producing comparable results. They have also adapted Chill to learn parsers that generate database queries rather than syntactic parse trees. Finally, Thompson (1995) has developed another system, used in conjunction with Chill, that acquires meanings for words used in training sentences. Overall, their research program makes stronger contact with core problems and concepts of AI than either the machine learning or the natural language communities.

Machine Learning for Planning

Another core area of AI involves planning and problem solving. The 1980's saw considerable work on this topic within the machine learning community, some of it collected in Minton (1993), but interest in the area has declined in recent years. However, Veloso and her colleagues (e.g., Veloso & Carbonell, 1993) have continued an active research program on learning in planning, carried out within the Prodigy architecture. This work has focused on derivational analogy, an approach to case-based learning that indexes plan components by the reasons they were selected and uses those reasons to decide whether to reuse them on new problems.

Unlike most research on the acquisition of plan knowledge, Veloso has tested her ideas on large problems from a difficult logistics domain. The primary measure of performance has been the overall planning time, but more recent work has also examined ability to improve the quality of the generated plans. Like most work on plan learning, Veloso's approach uses knowledge-guided means-ends analysis to produce totally-ordered plans. This stands in sharp contrast with the wisdom of the AI planning community in favor of partial-order planners. This difference has led to experimental studies (Veloso & Blythe, 1994; Kambhampati, Ihrig, & Srivastava, 1996) that have clarified the strengths of these two approaches and their underlying relationships.

Machine Learning for Physical Control

An intelligent agent must do more than perceive its surroundings and plan its actions; it must also execute those actions in the world. Although there has been some learning work within the robotics community, this has focused mainly on constructing maps from sensor data (e.g., Pierce & Kuipers, 1994). Within machine learning, some researchers have made progress on acquiring models for robot actions, but often by transforming this task into a one-step regression problem (e.g., Moore, 1990). Reinforcement learning claims to address the acquisition of action strategies, but most work on that topic has dealt with learning discrete actions in 'grid worlds', with little evidence that it generalizes to more realistic domains (e.g., Sutton, 1988).

However, work on reinforcement learning by Grefenstette and his colleagues (Grefenstette, 1988; Grefenstette, Ramsey, & Schultz, 1990; Grefenstette & Schultz, 1994) differs from this trend in some important respects. Rather than dealing with discrete domains, they have tested the ability of their Samuel system to learn control schemes in simulated continuous domains, using tasks like evading a heat-seeking missile and navigating through a mine field. They have also tested its methods on navigation tasks for physical robots, such as obstacle avoidance and following behaviors. And they have examined the role of control learning in both single-agent and multi-agent settings.

Moreover, Grefenstette et al.'s approach differs from mainstream methods for reinforcement learning, which estimate an expected reward for state-action pairs and then generalize across different states, thus treating each state as an independent classification problem. In contrast, Samuel searches through a space of control strategies, in which each strategy consists of condition-action rules that may interact in complex ways. The system uses genetic algorithms to generate new candidate strategies and directs search by evaluating the behavior of each strategy on a set of training problems, measuring the overall behavior rather than focusing on individual rules. This concern with explicit control programs, and the domains on which it has been tested, makes this work more akin to traditional AI research than that on temporal-difference methods.

Machine Learning for Computer Vision

Yet another example comes from the area of computer vision, one of the earliest areas to consciously separate from mainstream AI. Learning research has become quite common within the vision community in the past few years, as reported by Bowyer et al. (1994), but most work on this topic has more in common with pattern recognition than artificial intelligence. That is, the typical approach treats vision as a one-step classification problem, a view that has gained increasing adherents, rather than as a multi-step process of image understanding, which was the traditional AI stance.

One researcher who has not abandoned the image-understanding paradigm, and who has studied learning within this framework, is Draper (1993). Building on a multi-level vision system developed by Hanson, Riseman, and colleagues, he viewed image understanding as problem-space search in which the operators transformed a raw image into a structured description. This suggested the adaptation of techniques from learning search heuristics to determine the conditions under which to select each visual operator. As in Zelle and Mooney's work, he used solution paths found through search to generate positive and negative examples for use in training. However, rather than aiming at learning legal 'parses', Draper's concern was improving the efficiency of the image-understanding process.

More recently, Draper (1996) has shifted to using temporal-difference methods for reinforcement learning. The idea here is that a particular sequence of visual operators is not right or wrong, but only better or worse than some other sequence. Thus, instead of learning all-or-none conditions for each operator, his new approach combines backpropagation with temporal-difference calculations to induce an evaluation function from delayed rewards (the scores for parsed images). His system, which he has tested on both ground-level imagery and aerial photographs, then uses this learned metric to direct best-first search through the space of processed images. This multi-level approach differs sharply both from most work in visual learning and most work in the machine learning community.

Adaptive Advisory Systems

Although the central dream of AI is to construct an autonomous intelligent agent, a more realistic near-term goal is to create intelligent aids for humans. In fact, much of the 1980's excitement about 'expert systems' has now been replaced by excitement about 'advisory systems', which seem especially appropriate for many tasks that arise on the World Wide Web. As we have noted elsewhere (Langley & Simon, 1995), such advisory systems provide ideal environments for learning, since each decision by the user to accept or override a system recommendation generates a training case that it can use to improve future behavior.

Of course, we can divide advisory systems into those for simple prediction and those for more complex tasks, just as we can for autonomous systems. Many of the adaptive Web 'agents' under development, including most recommendation systems and information retrieval systems, fit the former bill and need not concern us here. But some work in this area has dealt with multi-step problems that, although giving humans ultimate authority, must still deal with many of the core issues in artificial intelligence.

Perhaps the best example of such work comes from Schlimmer and Hermen (1993, 1994), who have championed the idea of self-customizing software that alters itself through interaction with its users. Their adaptive system for form filling learns rules that predict default values for some fields based on the values of earlier ones, and thus greatly reduce keystrokes. Their advisory system system for note taking learns a grammar that predicts the order in which users will enter information, thus helping them organize their thoughts. Schlimmer and Hermen's research has focused on these and other multi-step tasks that feel much more like AI problems than information retrieval or product recommendation.

New Problems and New Methods

Clearly, we feel that machine learning researchers would benefit from focusing more effort on the core problems of AI, but we have seen that some scientists in this community have continued to work on natural language, planning, control, and computer vision. Yet the role of learning in some other areas of AI has been ignored almost entirely both by machine learning researchers and by scientists in those areas.

Perhaps the most glaring omission of this sort concerns tasks that are typically formulated as constraint-satisfaction problems. For instance, scheduling has received increasing attention in both experimental and applied AI circles, and there exists a clear role for learning heuristics to constrain search on large problems. But there has been almost no research within either the machine learning or scheduling communities along these lines. One exception is the work by Eskey and Zweben (1990), which drew on ideas from earlier approaches to learning symbolic heuristics for problem solving. More recently, Zhang and Dietterich (1995) have adapted methods from reinforcement learning to acquire evaluation functions for directing search. Both tested their methods on complex problems from a NASA scheduling domain with some success.

Configuration, which is also usually viewed in terms of constraint satisfaction, is another likely area for the use of learning, but again there has been almost no work on this topic by researchers in either community. The literature includes a few exceptions: Schlimmer (1991) has used rule-induction methods to learn constraints for error checking in computer configuration, Henessy and Hinkle (1992) have used case-based methods to learn acceptable layouts for convection ovens, and Reich and Fenves (1991) have adapted conceptual clustering to characterize workable bridge designs. But given the growing applied interest in configuration tasks, the omission seems almost incredible. Day (1992) and Minton (1996) have used learning in other constraint-satisfaction tasks, but again this work is the exception rather than the rule.

A broader issue, that cuts across many areas, concerns the need for learning at multiple levels. Natural language processing, image understanding, and plan generation, at least in their full form, all require computation at different levels of aggregation, yet most learning research on these topics deals with only one level. For example, AI approaches to vision posit levels for extracting edges, surfaces, primitive objects, and complex objects. We need to develop new frameworks that let knowledge acquired at one level of aggregation support learning at higher levels. One of the best-established results in machine learning is that appropriate knowledge can aid the learning process, yet this knowledge must itself be acquired in some fashion. Developing such multi-level learning systems will require identifying places where knowledge comes into play at each level and finding sources of training data that can drive learning.[3] This in turn will require a relatively complete framework for natural language, vision, or planning, and enough familiarity with those frameworks to embed learning.

These new tasks suggest the need for new types of learning methods, since it seems unlikely that supervised techniques or temporal-difference methods will be able to handle them alone. The cognitive psychology literature suggests one key: expert performance almost always shadows expert memory, which is typically explained through the acquisition of *chunks* (Miller, 1956). Theoretically, chunks are hierarchical structures made up from other chunks that, ultimately, are grounded in primitive percepts or actions. Within machine learning, the notion of macro-operators (e.g., Iba, 1988) come closest to the psychological notion of chunks, but work on this topic has nearly died out in recent years.[4]

However, some research on language learning employs methods for creating more flexible hierarchical structures. Wolff (1980), and more recently Stolcke and Omohundro (1994) and Langley (1995, Chapter 9) report algorithms that construct and merge nonterminal symbols for context-free grammars, giving rewrite rules at increasing levels of aggregation. Although they have tested this idea only on language-learning tasks, the basic idea should also prove useful in learning problem-reduction rules for planning and in acquiring complex motor skills. Such techniques have a quite different feel from methods for supervised concept induction and temporal-difference learning.

Another issue is that nearly all research in machine learning focuses on *induction*, in which multiple experiences lead to general rules or laws. Yet some forms of human learning and discovery instead involve increased understanding of single observations or events. The aim here is to construct a model that explains some observation in terms of existing domain knowledge. For example, upon observing the first pulsar, astronomers created a model that drew on their knowledge of stars and physics to explain its anomalous but regular behavior. There has been some work along these lines, such as that by Shrager (1987) on mechanical devices and Valdes-Perez (1992) on chemical reactions, but these are exceptions in a field obsessed with induction rather than learning in its broader senses.

Prospects for the Future

The goal seems clear: we must find ways to encourage research in machine learning that makes closer contact with the central goals of AI and, more broadly, achieve the same effect in other subareas. The path to this goal is much less obvious, but we can consider some actions that should take us in the direction of a more unified discipline.

The problem is partly technical in nature. The AI community has divided itself for a good reason – intelligence is a complex phenomenon and the strategy of divide and conquer make eminent sense in such situations. In this case, AI has split into subfields that study basic components of intelligence, such as planning and language, and basic issues that cut across these components, such as representation and learning.

Yet we know from the planning literature that there is often more than one way to decompose a difficult problem. A viable alternative would be for AI to partition itself into subfields that focus on constrained domains that require integration of the traditional areas. Recent progress on mobile robotics, believable agents that operate in simulated worlds, and intelligent advisory systems have involved such integration efforts, and their grouping under the 'agents' banner is no accident. However limited in scope, these systems differ from most AI systems and attract attention because they actually *do* something in nontrivial environments.

Another aspect of the problem is sociological. For various reasons, scientific fields have a natural tendency to fragment into more specialized areas, and only when there exist pressures against such division is the process slowed or reversed. In principle, we could alter this trend by changing the reward structure for AI research. For instance, we might bias paper and speaker selection against those who obtain quick results on simple tasks in favor of those who place their work in a broader context, even though their results are more preliminary. However, this would be difficult to implement, as many reviewers are recent PhDs trained to believe in their narrow paradigm.

[3] Laird, Pearson, and Huffman (1995) report a rare example of such work, in which they examine learning at the reactive, deliberate, and reflective levels in a simulated physical agent.

[4] Research on chunk creation within the Soar framework (Laird, Rosenbloom, & Newell, 1986) is still active, but their use of 'chunk' differs from the standard sense in psychology.

Thus, we also need to change our educational process. Graduate programs should begin to train AI students not just in machine learning or natural language, but also in knowledge representation, planning, and vision. New PhDs in artificial intelligence often share more knowledge about peripheral topics like automata theory than they do about core issues in the nature of intelligence. At the very least, students should have training across a number of areas, but we might go farther and replace the current courses and texts with ones that encourage integration.

Moreover, education is a continuing process, and we can complement broader formal training with informal efforts. Our reviews of submitted papers and our questions after talks can encourage researchers to address broader issues, we can recruit invited talks and tutorials that explore relations among AI's subfields, and we can even design conference sessions that cross traditional boundaries. Only in these ways can we raise consciousness about the need to move beyond existing paradigms toward a more unified approach to intelligent behavior.

Acknowledgements

Thanks to David Moriarty, Seth Rogers, and Stephanie Sage for providing useful comments on an earlier draft of this paper, and to the researchers reported earlier for maintaining their broad research agendas.

References

Bowyer, K. W., Hall, L. O., Langley, P., Bhanu, B., & Draper, B. A. (1994). Report of the AAAI Fall Symposium on Machine Learning and Computer Vision: What, Why and How. *Proceedings of the Image Understanding Workshop* (pp. 727–731). Monterrey, CA. Morgan Kaufmann.

Day, D. (1992). Acquiring search heuristics automatically for constraint-based planning and scheduling. *Proceedings of the First International Conference on AI Planning Systems* (pp. 45–51). College Park, MD: Morgan Kaufmann.

Draper, B. (1993). Learning from the schema learning system. *Working Notes of the AAAI Fall Symposium on Machine Learning in Computer Vision* (pp. 75–79). Raleigh, NC: AAAI Press.

Draper, B. (1996). Learning grouping strategies for 2D and 3D object recognition. *Proceedings of the Image Understanding Workshop* (pp. 1447–1454). Palm Springs, CA: Morgan Kaufmann.

Eskey, M., & Zweben, M. (1990). Learning search control for constraint-based scheduling. *Proceedings of the Eighth National Conference on Artificial Intelligence* (pp. 908–915). Boston: AAAI Press.

Grefenstette, J. J. (1988). Credit assignment in rule discovery systems based on genetic algorithms. *Machine Learning*, *3*, 225–245.

Grefenstette, J. J., Ramsey, C. L., & Schultz, A. C. (1990). Learning sequential decision rules using simulation models and competition. *Machine Learning*, *5*, 355–381.

Grefenstette, J. J., & Schultz, A. (1994). An evolutionary approach to learning in robots. *Proceedings of the Machine Learning Workshop on Robot Learning*. New Brunswick, NJ.

Hermens, L. A., & Schlimmer, J. C. (1994). A machine-learning apprentice for the completion of repetitive forms. *IEEE Expert*, *9*, 28–33.

Hennessy, D., & Hinkle, D. (1992). Applying case-based reasoning to autoclave loading. *IEEE Expert*, *7*, October, 21–26.

Iba, G. A. (1989). A heuristic approach to the discovery of macro-operators. *Machine Learning*, *3*, 285–317.

Kambhampati, S., Ihrig, I., & Srivastava, B. (1996). A candidate set based analysis of subgoal interaction in conjunctive goal planning. *Proceedings of the Third International Conference on AI Planning Systems* (pp. 125–133). Edinburgh: Morgan Kaufmann.

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

Laird, J. E., Pearson, D. J., & Huffman, S. B. (1996). Knowledge-directed adaptation in multi-level agents. *Proceedings of the AAAI Workshop on Intelligent Adaptive Agents*. Portland: AAAI Press.

Langley, P. (1995). *Elements of machine learning*. San Francisco: Morgan Kaufmann.

Langley, P., & Simon, H. A. (1981). The central role of learning in cognition. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum.

Langley, P., & Simon, H. A. (1995). Applications of machine learning and rule induction. *Communications of the ACM*, *38*, November, 55–64.

Miller, G. A. (1956). The magical number seven, plus or minus two. *Psychological Review*, *63*, 81–97.

Minton, S. (Ed.). (1993). *Machine learning methods for planning*. San Francisco, CA: Morgan Kaufmann.

Minton, S. (1996). Automatically configuring constraint satisfaction programs: A case study. *Constraints*, *1*, 7–43.

Moore, A. W. (1990). Acquisition of dynamic control knowledge for a robot manipulator. *Proceedings of the Seventh International Conference on Machine Learning* (pp. 244–252). Austin: Morgan Kaufmann.

Pierce, D., & Kuipers, B. (1994). Learning to explore and build maps. *Proceedings of the Twelfth National*

Conference on Artificial Intelligence (pp. 1264–1271). Seattle: AAAI Press.

Reich, Y., & Fenves, S. J. (1991). The formation and use of abstract concepts in design. In D. H. Fisher, M. J. Pazzani, & P. Langley (Eds.), *Concept formation: Knowledge and experience in unsupervised learning*. San Francisco, CA: Morgan Kaufmann.

Schlimmer, J. C. (1991). Learning meta knowledge for database checking. *Proceedings of the Tenth National Conference on Artificial Intelligence* (pp. 335–340). Anaheim, CA: AAAI Press.

Schlimmer, J. C., & Hermens, L. A. (1993). Software agents: Completing patterns and constructing interfaces. *Journal of Artificial Intelligence Research*, *1*, 61–89.

Shrager, J. (1987). Theory change via view application in instructionless learning. *Machine Learning*, *2*, 247–276.

Sleeman, D., Langley, P., & Mitchell, T. (1982). Learning from solution paths: An approach to the credit assignment problem. *AI Magazine*, *3*, 48–52.

Stolcke, A., & Omohundro, S. (1994). Inducing probabilistic grammars by Bayesian model merging. *Proceedings of the Second International Conference on Grammatical Inference and Applications* (pp. 106–118). Alicante, Spain: Springer-Verlag.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, *3*, 9–44.

Thompson, C. A. (1995). Acquisition of a lexicon from semantic representations of sentences. *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics* (pp. 335–337).

Valdes-Perez, R. E. (1992). Theory-driven discovery of reaction pathways in the Mechem system. *Proceedings of the Tenth National Conference on Artificial Intelligence* (pp. 63–69). San Jose: AAAI Press.

Veloso, M., & Blythe, J. (1994). Linkability: Examining causal link commitments in partial-order planning. *Proceedings of the Second International Conference on AI Planning Systems* (pp. 170–175). Chicago: Morgan Kaufmann.

Veloso, M. M., & Carbonell, J. G. (1993). Derivational analogy in Prodigy: Automating case acquisition, storage, and utilization. *Machine Learning*, *10*, 249–278.

Wolff, J. G. (1980). Language acquisition and the discovery of phrase structure. *Language and Speech*, *23*, 255–269.

Zelle, J. M., & Mooney, R. J. (1993). Learning semantic grammars with constructive inductive logic programming. *Proceedings of the Eleventh National Conference on Artificial Intelligence* (pp. 817–822). Washington, DC: AAAI Press.

Zelle, J. M., & Mooney, R. J. (1996). Learning to parse database queries using inductive logic programming. *Proceedings of the Thirteenth National Conference on Artificial Intelligence* (pp. 1050–1055). Portland: AAAI Press.

Zhang, W., & Dietterich, T. G. (1995). A reinforcement learning approach to job-shop scheduling. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (pp. 1114–1120). Montreal: Morgan Kaufmann.