Not the Path to Perdition: The Utility of Similarity-Based Learning

Michael Lebowitz[1]

Department of Computer Science -- Columbia University
New York, NY 10027

Abstract

A large portion of the research in machine learning has involved a paradigm of comparing many examples and analyzing them in terms of similarities and differences, assuming that the resulting generalizations will have applicability to new examples. While such research has been very successful, it is by no means obvious why similarity-based generalizations should be useful, since they may simply reflect coincidences. Proponents of explanation-based learning, a new, knowledge-intensive method of examining single examples to derive generalizations based on underlying causal models, could contend that their methods are more fundamentally grounded, and that there is no need to look for similarities across examples. In this paper, we present the issues, and then show why similarity-based methods *are* important. We present four reasons why robust machine learning must involve the integration of similarity-based and explanation-based methods. We argue that: 1) it may not always be practical or even possible to determine a causal explanation; 2) similarity usually implies causality; 3) similarity-based generalizations can be refined over time; 4) similarity-based and explanation-based methods complement each other in important ways.

1 Introduction

Until recently, machine learning has focused upon a single paradigm -- the generalization of concepts through the comparison of examples. The assumption has been made, though often tacitly, that the generalization of similarities will lead to concepts that can be applied in other contexts. Despite its ubiquity there is one real problem with this paradigm: there is no obvious reason why the underlying assumption should hold. In other fields people have called into doubt the utility of noticing similarities in the world and assuming them to be important. Naturalist Stephen Jay Gould, in discussing the nature of scientific discovery comments that:

> The human mind delights in finding pattern -- so much so that we often mistake coincidence or forced analogy for profound meaning. No other habit of thought lies so deeply within the soul of a small creature trying to make sense of a complex world not constructed for it.
>
> 'Into this Universe, and why not knowing // Nor whence, like water willy-nilly flowing' as the *Rubaiyat* says. No other habit of thought stands so doggedly in the way of any forthright attempt to understand some of the world's most essential aspects -- the tortuous paths of history, the unpredictability of complex systems, and the lack of causal connection among events superficially similar.
>
> Numerical coincidence is a common path to intellectual perdition in our quest for meaning. [Gould 84]

Further doubt has been cast upon the use of similarity-based learning by a new methodology that has been developed in the last few years: the extensive application of knowledge to single examples to determine the underlying mechanism behind an example, and the use of this causal explanation to derive generalized concepts. By learning from single examples, this knowledge-based approach calls into question the necessity of similarity-based approaches.

Despite Gould's warning and the recent successes of explanation-based methods, learning methods that concentrate on seeking out coincidences have had remarkable success across a variety of tasks. Furthermore, as Gould implies above, people (and other creatures) do seem to be optimized for such learning. Given this evidence, it worth trying to explain why such methods work. In this paper we will explain why similarity-based learning not only works, but is a crucial part of learning.

2 EBL and SBL

Considerable research has been done involving *similarity-based learning* (SBL). [Winston 72; Winston 80; Michalski 80; Michalski 83; Dietterich and Michalski 86; Lebowitz 83; Lebowitz 86a] are just a few examples. (See also, [Michalski et al. 83; Michalski et al. 86].) While there are many variations to such learning research, the basic idea is that a program takes a number of examples, compares them in terms of similarities and differences, and creates a generalized description by abstracting out similarities. A program given descriptions of Columbia University and Yale University and told that they were Ivy League universities and that the University of Massachusetts was not would define "Ivy League university" in terms of the properties that the first two examples had and that the third did not -- e.g., as being private, expensive and old. Similarity-based learning has been studied for cases where the input is specially prepared by a teacher; for unprepared input; where there are only positive examples; where there are both positive and negative examples; for a few examples; for many examples; for determining only a single concept at a time; and for determining multiple concepts. In a practical sense, SBL programs have learned by comparing examples more or less syntactically, using little "high level" knowledge of their domains (other than in deciding how to represent each example initially).

Explanation-based learning (EBL), in contrast, views learning as a knowledge-intensive activity, much like other tasks in Artificial Intelligence. [DeJong 86; Ellman 85; Mitchell 83a; Mostow 83; Minton 84; Silver 86] are a few examples of explanation-based learning research. (See also [Michalski et al. 86].) An EBL program takes a single example, builds up an explanation of how the various components relate to each other at a low level of detail by using traditional AI understanding or planning methods, and then generalizes the properties of various components of the example so long as the explanation remains valid. What is left is then viewed as a generalized description of the example that can be applied in understanding further examples. This kind of learning is tremendously useful, as it allows generalized concepts to be determined on the basis of a single example. On the other hand, the building and analysis of explanations does require extremely detailed knowledge of the domain (which may minimize the need to learn). In addition, virtually all current EBL work is in the "perfect learner" paradigm that assumes that all input is noise-free and fits the correct final generalization.

[1]This research was supported in part by the Defense Advanced Research Projects Agency under contract N00039-84-C-0165 and in part by the United States Army Research Institute under contract MDA903-85-0103. Comments by Kathy McKeown on an earlier draft of this paper were quite useful.

It is important to make clear here exactly the sense in which EBL is concept learning. It might be contended that all that is being done is the application of pre-existing information to a problem, unlike SBL, which is clearly a form of inductive learning. The key is in the generalization phase, where the EBL learner loosens constraints on its representation and determines whether the explanation that it has built up still holds. This generalized concept can then serve as a form of compiled knowledge that simplifies the processing of later input. This may be a way to learn structures such as frames [Minsky 75] and scripts [Schank and Abelson 77]. The view of using EBL to produce knowledge structures that make later processing more efficient has been called *operationalization* [Mostow 83]. Even though it might in some sense be possible to understand later examples just using low-level rules, realistically it is crucial to have a set of knowledge structures at various levels of complexity.

3 The goal of learning

It does not make sense to consider learning in isolation from other elements of intelligent processing. While certain aspects of learning may not be in service of an immediate goal (e.g., curiosity), at some point there must be a task involved to make use of what is learned. In general, the idea is for an organism or program to be able to carry out a task better (either be able to do more examples or do examples more efficiently) than it did before learning. It is particularly important to keep in mind the task nature of learning when considering concept learning, which has often been studied without regard to the future utility of the concepts created.

For most tasks that people or intelligent programs will carry out, the most obvious way to be able to improve performance is to attempt to develop a *causal model* that explains how elements of the domain work. Such a model will allow the learner to *predict* what is likely to happen in later situations, which will clearly be useful. The model will allow the learner to *understand* further input. Although we will consider later whether it is possible in all domains, the construction of a causal model is clearly a worthy goal in learning. [Schank 75; Schank 84] present reasons for constructing such models even in domains with incomplete models. Explanation-based learning methods strike directly at the problem of creating causal models. Similarity-based methods do not, but yet seem to lead to useful generalizations. This leads us to the central mystery of this paper.

4 The puzzle

Having decided that the construction of a causal model for a domain is important, or perhaps even crucial, as part of learning, we are left with the key question, "Is there any role for similarity-based learning in a full learning model, and if so, why?" Even if we assume that there must be something to SBL, since, after all, so many people have worked on it with impressive results, we must ask why it works; why it helps a learner perform better. That generalizations from explanation-based learning are valid and useful makes sense intuitively, since they are derived from causal analyses. Similarity-based generalizations could just be the result of the coincidences that arise in a complex world.

Note that similarity-based learning is not merely an artifact of researchers in machine learning. As pointed out in the Gould quote above, people delight in noticing similarities in disparate situations. Indeed, in many ways human processing seems to be optimized for such learning. An anecdotal example immediately comes to mind: On the Eastern Air Shuttle between New York and Boston, passengers are given a sequence number for boarding. On one roundtrip, I received the same sequence number going in each direction. I noticed the similarity immediately, even though the first number was not in front of me when I received the second, despite the apparent irrelevance of the coincidence to my performance on later shuttle trips. Virtually everyone has experienced, and noticed, similar coincidences. When nature provides such a powerful cognitive mechanism, there always seems to be a good reason. We will see shortly why the recognition of similarities is important, though, to reiterate, the utility is *not* obvious and should not simply be assumed by SBL researchers.

5 A similarity-based learning program

We can most easily look at the utility of SBL in the context of a specific learning program. UNIMEM [Lebowitz 82; Lebowitz 86a; Lebowitz 86b] takes examples represented as sets of features (essentially property/value pairs) and automatically builds up a generalization hierarchy using similarity-based methods. It is not told in advance which examples to compare or concepts to form, but instead learns by observation. One domain on which we have tested UNIMEM involves data about universities that was collected from students in an Artificial Intelligence class at Columbia.[2]

Figure 1 shows the information used by UNIMEM for two universities, Columbia and Carnegie-Mellon. Each university is represented by a set of triples that describe features of the university, the first two providing a property name and the third its value. So, Columbia is in New York State while Carnegie-Mellon is in Pennsylvania. Both are urban and private and Columbia has a 7/3 male/female ratio compared to Carnegie-Mellon's 6/4. Some features, like quality of life, involve arbitrary numeric scales.

FEATURE:		COLUMBIA:	CMU:
STATE	VALUE	NEW-YORK	PENNSYLVANIA
LOCATION	VALUE	URBAN	URBAN
CONTROL	VALUE	PRIVATE	PRIVATE
MALE:FEMALE	VALUE	RATIO:7:3	RATIO:6:4
NO-OF-STUDENTS	VALUE	THOUS:5-	THOUS:5-
STUDENT:FACULTY	VALUE	RATIO:9:1	RATIO:10:1
SAT	VERBAL	625	600
	MATH	650	650
EXPENSES	VALUE	THOUS$:10+	THOUS$:10+
%-FINANCIAL-AID	VALUE	60	70
NO-APPLICANTS	VALUE	THOUS:4-7	THOUS:4-7
%-ADMITTANCE	VALUE	30	40
%-ENROLLED	VALUE	50	50
ACADEMICS	SCALE:1-5	5	4
SOCIAL	SCALE:1-5	3	3
QUALITY-OF-LIFE	SCALE:1-5	3	3
ACAD-EMPHASIS	VALUE	LIB-ARTS	ENGINEERING

Figure 1: Information about two universities

The first question we have to address concerning the examples in Figure 1 is precisely what it means to "understand" them, or to learn from them. While the exact nature of understanding would depend on the ultimate task that we had in mind, presumably what a person or system learning from these examples would be after is a causal model that relates the various features to each other.

As an example, in understanding Figure 1 we might wish to know how the fact that both universities are private relates to the fact that they are both expensive or why Carnegie-Mellon offers financial aid to more people. A causal model that answers questions of this sort would be extremely useful for almost any task involving universities. Typical of the causation that we would look for is, for example, that private universities get less government support and hence have to raise more money through tuition. (At least that is how private universities explain it!) Similarly, a model

[2]Other domains UNIMEM has been tested on include: information about states of the United States, Congressional voting records, software evaluations, biological data, football plays, universities, and terrorism stories.

might indicate that Carnegie-Mellon's emphasis on engineering leads to the acceptance of more students who need financial aid. Notice, however, that it will certainly not be possible to build a complete causal model solely from the information in Figure 1, but will require additional domain knowledge.

An EBL program would create a low-level causal model of a university using whatever methods were available and then would use the model to develop a generalized concept. For example, it might decide that the Columbia explanation could be generalized by removing the requirement of being in New York State and by allowing the numeric values to vary within ranges, if none of these changes would affect the underlying explanation. It might be, however, that the liberal arts emphasis is crucial for some aspect of the explanation. In any case, by relaxing constraints in the representation, an EBL program would develop, using a single, causally motivated example, a generalized concept that ought to apply to a wide range of situations.

Let us now compare the desired causal explanation with the kind of generalization made using similarity-based methods. Figure 2 shows the generalization that is made by UNIMEM, GND1, from the two university representations in Figure 1.[3] We see in Figure 2 that UNIMEM has generalized Columbia and Carnegie-Mellon by retaining the features that have identical values (like social level and quality of life), averaging feature values that are close (such as SAT verbal score) and eliminating features that are substantially different, such as the state where the university is located and the percentage of financial aid.[4] The resulting set of features can be viewed as a generalization of the two examples, as it describes both of them, as well as, presumably, other universities that differ in other features.

```
GND1
    SOCIAL              SCALE:1-5     3
    QUALITY-OF-LIFE     SCALE:1-5     3
    LOCATION            VALUE         URBAN
    CONTROL             VALUE         PRIVATE
    NO-OF-STUDENTS      VALUE         THOUS:5-
    STUDENT:FACULTY     VALUE         RATIO:9:1
    SAT                 MATH          650
    SAT                 VERBAL        612.5
    EXPENSES            VALUE         THOUS$:10+
    NO-APPLICANTS       VALUE         THOUS:4-7
    %-ENROLLED          VALUE         50
    [CARNEGIE-MELLON COLUMBIA]
```

Figure 2: Generalizing Columbia and Carnegie-Mellon

What would the generalization in Figure 2 be used for once it had been made? Presumably it would be used in processing information about other universities. If we identified a situation where GND1 was thought to be relevant, we would assume that any of its features that were not known would indeed be present. The assumption is made by all similarity-based learning programs, including UNIMEM, that they have created usable concepts from which default values may be inherited.

We can now state our problem quite clearly in terms of this example: *What reason do we have to believe that a new example that fits part of the generalization of Columbia and Carnegie-Mellon will fit the rest?* With explanation-based methods we at least have the underlying causal model as justification for believing the generalization. But what is the support of similarity-based learning?

6 Elements of an answer

There are four main elements to our explanation as to why SBL produces generalized concepts that can be profitably applied to other problems and why it should be so used:

- While the goal of learning is indeed a causal model, it is often not possible to determine underlying causality and even where it is possible it may not be practical.
- Similarity usually implies causality and is much easier to determine.
- There are ways to *refine* generalizations to mitigate the effects of coincidence.
- Explanation-based and similarity-based methods complement each other in crucial ways.

6.1 Causality cannot always be determined

In order to achieve their impressive results, the EBL methods that have been developed to date assume that a complete model of a domain is available and thus a full causal explanation can be constructed. In addition, it is assumed that it is always computationally feasible to determine the explanation of any given example. While these assumptions may be acceptable for some learning tasks, they do not appear reasonable for situations where we are dealing with noisy, complex, uncertain data -- characteristics of most real-world problems. It is also unreasonable to expect to have a complete domain model available for a new domain that we are just beginning to explore. Even in our university example, it is hard imagine all the information being available to build a complete model.

Most EBL work has not addressed these issues. Some of the domains used, like integration problems [Mitchell 83a], logic circuits [Mitchell 83b; Ellman 85] or chess games [Minton 84] do indeed have complete domain models and the examples used are small enough for the explanation construction to be tractable. Even in a domain such as the news stories of [DeJong 86], the assumption is made, perhaps less validly, that it is always possible to build up a complete explanation.

In domains where a detailed explanation cannot reasonably be constructed, a learner can only rely on similarity-based methods. By looking for similarities it is at least possible for the learner to bring some regularity to its knowledge base. The noticing of co-occurrence is possible even the absence of a complete domain model. Further, much research, including our own, has shown that SBL can be done efficiently in a variety of different problem situations. In the university example of Section 5, UNIMEM was able to come up with a variety of similarity-based generalizations with minimal domain information. Further, as we noted above, people seem to be optimized for SBL.

6.2 Similarity usually implies causality

The regularity that is detected using SBL is not worthwhile if it cannot be used to help cope with further examples. Such help is not likely if there is no connection between the similarities and the underlying causal explanation. Fortunately, such a connection will usually exist.

Put as simply as is possible, similarities among examples usually occur because of some underlying causal mechanism. Clearly if there is a consistent mechanism, it will produce consistent results that can be observed as similarities. While the infinite variety of the world will also produce many coincidental similarities, it is nonetheless true that among the observed similarities are the

[3] Actually, UNIMEM also had to decide that these two examples should even be compared and that they had a substantial amount in common before doing the actual generalization.

[4] Exactly what constitutes "substantially different" is a parameter of the program.

mechanisms that we desire.

So, in the Eastern Shuttle example used above, while it is almost certain that the duplicate seat numbers I received were coincidental, if there *was* a mechanism involving seat numbers (say the numbers were distributed in alphabetical order) it would manifest itself in this sort of coincidence. Similarly, in the university generalization GND1 (Figure 2), we indicated possible of mechanisms that would lead to the kind of expensive private school that is described.

Two recent examples illustrate how causal understanding frequently relates to similarity-based processing. The first involves scientific research, an attempt to understand a complex meteorological phenomenon, and the second an investigation into a mysterious crime.

In recent years weather researchers have been trying to explain a set of possibly related facts. Specifically: 1) the average temperature in 1981 was very high; 2) the El Chichon volcano erupted spectacularly in early 1982; 3) El Nino (a warm Pacific current) lasted an exceptionally long time starting in mid-1982; 4) there have been severe droughts in Africa since 1982.

One might expect researchers to immediately attempt to construct a causal model that explains all these phenomena. However, weather systems are extremely complex, and by no means fully understood. Author Gordon Williams, writing in *Atlantic*, discusses the attempt to gain understanding as follows: "How could so much human misery in Africa be caused by an errant current in the Pacific? *Records going back more than a century show that the worst African droughts often come in El Nino years.*" (Emphasis added.) Furthermore, Williams quotes climate analyst Eugene Rasmusson as saying, "It's disturbing because we don't understand the process" [Williams 86].

We can see clearly in this example that although the ultimate learning goal is a causal model, the construction of such a model is not immediately possible. So, researchers began by looking for correlations. However, they expect correlations to lead eventually to deeper understanding.

The second example involves investigators trying to determine how certain extra-strength Tylenol capsules became laced with poison. The *New York Times* of February 16, 1986 reported:

> Investigators tracing the routes of two bottles of Extra-Strength Tylenol containing cyanide-laced capsules have found that both were handled at the same distribution center in Pennsylvania two weeks apart last summer. Federal officials and the product's manufacturer said that the chance that the tainting occurred at the distribution facility was remote, but the finding prompted investigators to examine the possibility as part of their inquiry."

Again we have a case where a causal explanation is desired and yet there is not enough information available to construct one. So, the investigators began by looking for commonalities among the various poisoned capsules. When they found the distribution facility in common, that became an immediate possible contributor to the explanation. Although no final explanation had been discovered as this is written, it is clear that the explanation process attempted began with the noticing of similarities.

There is one further connection between noticing similarities and generating explanations that is worth making. This involves the idea of *predictability*. It turns out that the kinds of similarities that are noticed provide clues not only to what features should be involved in an explanation, but what the direction of causality might be (e.g., what causes what). As we have described elsewhere [Lebowitz 83; Lebowitz 86c], features that appear in just a few generalizations, which we call *predictive*, are the only ones that indicate a generalization's relevance to a given situation, and, further, are those likely to be the *causes* in an underlying explanation. This becomes clear when we realize that a feature present in many different situations cannot cause the other features in any single generalization, or it would cause the same features to appear in *all* the other generalizations that it is in.

In the weather example above, if we knew of many generalizations involving droughts, but only one with both warm currents and a volcano, then the volcano might cause the drought, but the drought could not cause the volcano. Of course, it may be that neither direction of causality is right, there being a common cause of both, but at least predictability provides a starting point.

The power of predictability is that it can be determined quite simply, basically as a byproduct of the normal SBL process. The various indexing schemes used in a generalization-based memory [Lebowitz 83; Lebowitz 86a] allow the simple counting of features in context. While there are many problems to be explored, particularly that of predictive combinations of features, the ability to know the likely initial causes when determining a mechanism is an important advantage of SBL. Further, even when no explanation can be found, the use of predictability often allows us to make predictions from a generalization at the correct moments, even without any deep understanding of the generalization

6.3 Refining generalizations

The third part of our explanation as to the utility of similarity-based learning is that generalizations, once made, are not immutable -- they can be *refined* in the light of later information. This means that the aspects of a generalizations that are due to coincidence can be removed. We have developed various techniques for doing this [Lebowitz 82] that work essentially by noticing during processing when various elements of a generalization are contradicted by new examples. If we remove the features that are frequently contradicted we can have a concept that is more widely applicable and contain meaningful information.

As an example of this, we will look again at our university generalization (Figure 2). Suppose that there were a wide range of universities with most of the features of GND1, but with different levels of social life. This contradiction of the social level value that was derived from the coincidental value that both Columbia and Carnegie-Mellon have might seem to invalidate the generalization. However, our refinement methods would allow UNIMEM (or a similar system) to remove this feature, leaving a more widely applicable generalization that describes high-quality private schools. In this way similarity-based methods can overcome some of the coincidences that might seem to require explanation-based methods. Notice, however, that UNIMEM makes this refinement without having any real idea of why it is doing so, other than the pragmatic rationale that it allows the generalization to fit more examples, but does not reduce it so much that it carries no information.

6.4 Integrated learning

The final element of our explanation for the importance of similarity-based methods lies in the need for an integrated approach employing both similarity-based and explanation-based approaches. This point is really a corollary of the relation between similarity and causality described in Section 6.2.

The basic idea is to use EBL primarily upon the generalizations that are found using SBL rather than trying to explain everything in sight. This drastically cuts down the search necessary for constructing an explanation, particularly in domains where we have very little specific knowledge and have to rely on general rules for the explanations. Basically, we use SBL as a bottom-up control on the top-down processing of EBL.

The "real world" weather and crime investigation examples in

Section 6.2 illustrate clearly how human problem solvers make use of this form of integrated learning -- trying to explain the coincidences that are noted, rather than explaining every element of a situation from scratch. We have described how a simple form of such integrated learning has been implemented for UNIMEM in [Lebowitz 86c]. For the university example in Figure 5, the main point is that we would only try to build up an explanation for the generalization GND1 (actually, the version of GND1 refined over time), and not the specific examples that made it up. Explaining the generalization is likely to be much easier than explaining the features of Columbia and Carnegie-Mellon and provide almost as much information.

7 Conclusion

We have shown in this paper a number of ways that similarity-based learning can contribute to the ultimate learning goal of building a coherent causal explanation of a situation. From this analysis it is not surprising that people seem to be optimized for noticing similarities, as such processing leads to the understanding that helps deal with the world. Our computer programs should be equally well equipped. Similarity-based learning is definitely not the path to perdition.

References

[DeJong 86] DeJong, G. F. An approach to learning from observation. In R. S. Michalski, J. G. Carbonell and T. M. Mitchell, Ed., *Machine Learning: An Artificial Intelligence Approach, Volume II*, Morgan Kaufmann, Los Altos, CA, 1986, pp. 571 - 590.

[Dietterich and Michalski 86] Dietterich, T. G. and Michalski, R. S. Learning to predict sequences. In R. S. Michalski, J. G. Carbonell and T. M. Mitchell, Ed., *Machine Learning: An Artificial Intelligence Approach, Volume II*, Morgan Kaufmann, Los Altos, CA, 1986, pp. 63 - 106.

[Ellman 85] Ellman, T. Generalizing logic circuit designs by analyzing proofs of correctness. Proceedings of the Ninth International Joint Conference on Artificial Intelligence, Los Angeles, 1985, pp. 643 - 646.

[Gould 84] Gould, S. J. "The rule of five." *Natural History 93*, 10, October 1984, pp. 14 - 23.

[Lebowitz 82] Lebowitz, M. "Correcting erroneous generalizations." *Cognition and Brain Theory 5*, 4, 1982, pp. 367 - 381.

[Lebowitz 83] Lebowitz, M. "Generalization from natural language text." *Cognitive Science 7*, 1, 1983, pp. 1 - 40.

[Lebowitz 86a] Lebowitz, M. Concept learning in a rich input domain: Generalization-Based Memory. In R. S. Michalski, J. G. Carbonell and T. M. Mitchell, Ed., *Machine Learning: An Artificial Intelligence Approach, Volume II*, Morgan Kaufmann, Los Altos, CA, 1986, pp. 193 - 214.

[Lebowitz 86b] Lebowitz, M. UNIMEM, a general learning system: An overview. Proceedings of ECAI-86, Brighton, England, 1986.

[Lebowitz 86c] Lebowitz, M. "Integrated learning: Controlling explanation." *Cognitive Science 10*, 2, 1986, pp. 219 - 240.

[Michalski 80] Michalski, R. S. "Pattern recognition as rule-guided inductive inference." *IEEE Transactions on Pattern Analysis and Machine Intelligence 2*, 4, 1980, pp. 349 - 361.

[Michalski 83] Michalski, R. S. "A theory and methodology of inductive learning." *Artificial Intelligence 20*, 1983, pp. 111 - 161.

[Michalski et al. 83] Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (Eds.). *Machine Learning, An Artificial Intelligence Approach.* Morgan Kaufmann, Los Altos, CA, 1983.

[Michalski et al. 86] Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (Eds.). *Machine Learning, An Artificial Intelligence Approach, Volume II.* Morgan Kaufmann, Los Altos, CA, 1986.

[Minsky 75] Minsky, M. A framework for representing knowledge. In P. H. Winston, Ed., *The Psychology of Computer Vision*, McGraw-Hill, New York, 1975.

[Minton 84] Minton, S. Constraint-based generalization. Proceedings of the Fourth National Conference on Artificial Intelligence, Austin, TX, 1984, pp. 251 - 254.

[Mitchell 83a] Mitchell, T. M. Learning and problem solving. Proceedings of the Eighth International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany, 1983, pp. 1139 - 1151.

[Mitchell 83b] Mitchell, T. M. An intelligent aid for circuit redesign. Proceedings of the Third National Conference on Artificial Intelligence, Washington, DC, 1983, pp. 274 - 278.

[Mostow 83] Mostow, J. Operationalizing advice: A problem-solving model. Proceedings of the 1983 International Machine Learning Workshop, Champaign-Urbana, Illinois, 1983, pp. 110 - 116.

[Schank 75] Schank, R. C. The structure of episodes in memory. In D. Bobrow and A. Collins, Ed., *Representation and Understanding: Studies in Cognitive Science*, Academic Press, New York, 1975, pp. 237 - 272.

[Schank 84] Schank, R. C. The Explanation Game. Technical Report 307, Yale University Department of Computer Science, New Haven, CT, 1984.

[Schank and Abelson 77] Schank, R. C. and Abelson, R. P. *Scripts, Plans, Goals and Understanding.* Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1977.

[Silver 86] Silver B. Precondition analysis: Learning control information. In R. S. Michalski, J. G. Carbonell and T. M. Mitchell, Ed., *Machine Learning: An Artificial Intelligence Approach, Volume II*, Morgan Kaufmann, Los Altos, CA, 1986, pp. 647 - 670.

[Williams 86] Williams, G. "The weather watchers." *Atlantic 257*, 1986, pp. 69 - 73.

[Winston 72] Winston, P. H. Learning structural descriptions from examples. In P. H. Winston, Ed., *The Psychology of Computer Vision*, McGraw-Hill, New York, 1972, pp. 157 - 209.

[Winston 80] Winston, P. H. "Learning and reasoning by analogy." *Communications of the ACM 23*, 1980, pp. 689 - 702.