The Future of Knowledge Representation
Extended Abstract

Ronald J. Brachman

AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

Abstract

Knowledge representation (KR) has traditionally been thought of as the heart of artificial intelligence. Anyone who has ever built an expert system, a natural language system—almost any AI system at all—has had to tackle the problem of representing its knowledge of the world. Despite it ubiquity, for most of AI's history KR has been a backstage activity. But in the 1980's it emerged as a field unto itself, with its own burgeoning literature. Along with this growth, the last decade has seen major changes in KR methodology, important technical contributions, and challenges to the basic assumptions of the field. I survey some of these developments, and then speculate about some of the equally interesting changes that appear on the horizon. I also look at some of the critical problems facing KR research in the near future, both technical and sociological.

Introduction

There is little doubt that the current level of achievement of artificial intelligence is largely due to the concept of a computational system capable of using an explicitly represented store of knowledge to allow it to reason about its goals, its environment, other agents, and itself. From its very beginnings, AI research has proceeded, often implicitly, on the assumption that the knowledge needed to get along in the world could be written down in some form, and then used as needed. The incarnations of this hypothesis have varied over the years, from McCarthy's "Advice Taker," to Newell and Simon's "Physical Symbol System Hypothesis," to the "Knowledge is Power" epithet of the expert system years, to Brian Smith's explicit statement of the "Knowledge Representation Hypothesis." While we have recently seen criticisms leveled at approaches that use explicit formalizations of knowledge, most of AI still rests securely on the foundation of knowledge representation (KR).

The last ten years have been remarkably productive for KR. The 1980's produced a number of critical new ideas and many important technical developments. The first dedicated international KR conference was held recently, and several of the more recent Computers and Thought Award winners have been from the mainstream of KR. Yet a closer look reveals that these are risky times as well. Some of the basic beliefs of the KR world are coming under increased scrutiny and doubt. Much of the work in the area has become technically obscure, accessible only to a few insiders. What "knowledge representation" means to KR insiders seems to be diverging from what it means to the rest of AI, and as a result, there is danger of the most important problems slipping through an ever-widening crack.

This article outlines a talk in which I attempt to look ahead and see what might be in store for knowledge representation in the next few years.[1] In it, I try to predict some of the trends that will set the pace for the next few years of research, and I will also outline some important issues that need to be addressed.

Before we can look ahead, it is important to determine where we have been and where we stand. To that end, I look back at the recent history of ideas in KR. First, I consider some of the general concerns of the field. Then, I touch on the style and state of work prior to 1980. Besides being an even ten years ago, 1980 was the year of the first NCAI; there, I reported on "Recent Developments in Representation Languages." Between that talk and the 1980 SIGART Newsletter Special Issue on KR (Brachman & Smith, 1980), we have a reasonable picture of the state of the art at that time.[2]

After giving a quick caricature of the early days, I outline some of the key developments of the last ten years. From these we can identify some broad trends that seem to be governing work as we move into the '90's. I attempt to project the likely course of some of those trends into the next few years. Finally, I look at some issues that may be critically important to the suc-

[1] Here I only briefly outline some of the topics to be addressed in the talk; this is not a transcript. The references are also not comprehensive.

[2] Also in 1980, Newell introduced the "knowledge level," and a special issue of *Artificial Intelligence* on "nonmonotonic reasoning" appeared. These were particularly significant harbingers of things to come in the 1980's.

cess of KR—and AI—over the next decade. I discuss several research problems and make some recommendations for movement in the field as a whole.

Of course, while it is easy to make predictions, it is quite another thing to be right. Much of this discussion should be taken with a rather large grain of salt. Furthermore, I cannot pretend to be either totally fair or complete. I have had a great deal of help, with ideas donated by many of the finest researchers in the field.[3] In the end, though, this represents only a personal opinion on a complicated and intertwined set of issues.

The Field

Interpretations of "knowledge representation" and its role in Artificial Intelligence vary widely, but at heart the idea is a simple and straightforward one: how do we impart knowledge of the world to a robot or other computational system so that, given an appropriate reasoning capacity, that knowledge can be used to allow the system to adapt to and exploit its environment? There are several basic issues:

- in what form is the knowledge to be expressed?
- how can the reasoning mechanism use a limited amount of knowledge to generate as much of the rest (i.e., implicit knowledge) as it needs?
- how can explicit and derived knowledge be used to influence the behavior of the system?
- what can the system do in the presence of incomplete or noisy information?
- how can the system actually reason in the face of potentially overwhelming search and inference complexity (e.g., NP-hard or undecidable problems)?

Since most AI systems have explicit knowledge bases of one form or another, the key issues of representation and reasoning are pervasive. Indeed, almost anything that involves the above issues can be considered work in "knowledge representation."

KR has always been principally concerned with the design of *forms* for expressing information, ranging from informal memory models to complex formal languages. Different KR systems may be better suited to different problems, but much of KR has proceeded under the banner of "general-purpose" KR languages and systems. In many cases, the interest is in how to represent a fragment of knowledge in a formal structure, without regard for how it will be used.

However, it is fairly widely held that it is virtually useless to consider a representation without considering the reasoning that is to be done with it. In much of KR the kind of reasoning that will be done is primary, and the structures used to represent the grist for the reasoning mill are secondary. "KR" now clearly stands for "Knowledge Representation and Reasoning." As a result, the study of KR is rooted in the study of logics,[4] where formal syntaxes of languages are accompanied by rules of inference and interpretations. This provides a standard for the correctness of an implemented "knowledge representation system"; without the logic, the only meaning of the representation language is the implementation itself. Many different types of reasoning are possible with the same syntax, and much of the concern in the field has been with "extra-logical" manipulations, such as belief revision.

In addition to the above, the field has aspects that provide it with a great richness, but also make it difficult to give a simple characterization. The variety of approaches also means that there are sometimes incomparable or conflicting goals in KR work. Much of it involves the form of knowledge, but increasingly, and appropriately, KR workers are dealing with content issues. Among the other research endeavors in KR we see at least the following:

- *foundational mathematics*—KR has developed its own repertoire of complex logics, and has used various theoretical tools to analyze the connections between them as well as the complexity of computing various functions over the representation structures.
- *cognitive science*—one part of KR research is primarily interested in the structure of human thought and its parameters, either for its own sake or in order to design and understand variations on the theme.
- *representing knowledge*—part of KR involves actually producing domain theories. Recent work has been concerned with more general "ontologies," axiomatizations that stretch across domains, the production of large, reusable knowledge bases, and issues of knowledge engineering and acquisition.
- *reasoning*—many forms of domain-independent, commonsense reasoning (e.g., nonmonotonic reasoning) are central to or overlap with KR.
- *technology*—KR researchers build systems to support a wide variety of AI implementations. KR systems can include knowledge base management facilities, interface tools, query languages, acquisition tools, etc.

So, while the basic idea of "knowledge representation" is simple, the field as a whole is complex and var-

[3] I am grateful to Phil Agre, Danny Bobrow, Alex Borgida, David Chapman, Ken Church, Robin Cohen, Jim des Rivieres, Jon Doyle, David Etherington, Mike Genesereth, Matt Ginsberg, Pat Hayes, Rick Hayes-Roth, David Israel, Henry Kautz, Wendy Lehnert, Hector Levesque, Bob MacGregor, Bill Mark, Drew McDermott, Bob Moore, Peter Patel-Schneider, Ramesh Patil, Ray Reiter, Chuck Rich, Len Schubert, Stu Shapiro, John Sowa, Mark Stefik, Bill Swartout, Peter Szolovits, Robert Wilensky, Yorick Wilks, and Brian Williams. Special thanks to Hector Levesque and Brian Williams for extensive discussions and assistance.

[4] This is not to say simply classical first-order predicate logic, but logics in general.

iegated. While the field has consolidated in some ways over the last few years, it has grown tremendously.

The Pioneering Days

It is hard to characterize an entire field over any length of time, but I will briefly try to give some general sense of the way things were in KR prior to 1980.[5]

In the early days of AI, KR was largely practiced as a subsidiary activity to more problem-specific tasks like natural language understanding. Prior to 1975 AI conferences did not have separate sessions for KR; instead KR papers showed up in sessions like "Computer Understanding" and "Natural Language: Semantics and Parsing." The roots of KR as we know it can be found in work on problem-solving at MIT and CMU in the late 1960's, and emerging work on semantic networks.

In the 1970's, as AI grew, KR's horizons expanded rapidly. The major body of work began in semantic nets, although there were occasional connections with formal logic, and numerous *ad hoc* formalisms. Many new representation languages were invented, although their intended scope and semantics were often unaddressed. It was assumed that KR systems were to be general-purpose support tools for virtually all of AI, and arguments about the superiority of one over another were often made with respect to anecdotal treatment of natural language examples. There were numerous arguments over what role—if any—mathematical logic should play in representing knowledge, and great discussion of whether "procedural" approaches were superior to "declarative" ones.[6] These were exciting, pioneering days, with a bit of a "land rush" feel: there was plenty of territory to be staked out, and as many ways of representing knowledge as there were people interested in the topic. Representative of the times was a panel on KR held at IJCAI-77, in which eight participants presented and argued the merits of their KR "hobby horses."

An important development was Minsky's frame paper (Minsky, 1975). Minsky's concern for more realistic commonsense reasoning, using prototypes and defaults, complex object descriptions, "differential diagnoses" of situations, etc., invigorated the KR community and led to the development of numerous frame representation systems (most of which did not really address Minsky's key insights).

In the mid-'70's, the lack of semantic accounts of representational formalisms and misunderstandings of the role of logic were growing concerns (e.g., see (Woods, 1975)). Towards the end of the decade, things began to sort themselves out, and the calls for treatment of semantics were increasingly heeded. In some cases, frame and semantic net systems were defined in terms of standard logics. In others, predicate logic was used directly as a representation medium. Generally, a more formal approach was beginning to take root.

By 1980, four representational paradigms seemed to predominate: semantic networks, frames, predicate logic, and production systems. A few people were experimenting with approaches combining two of these. Most approaches were still "general purpose," although there were several subareas of KR where reasoning of a specialized sort (e.g., qualitative physics) was primary and representation was tailored to the task. There were even some *de facto* standards emerging in the various frame and network representations, such as composite objects with slots, generalization hierarchies, inheritance of properties, and procedural attachment. Resource-limited processing and meta-description were popular topics. Qualitative physics and other reasoning areas had their own growing communities. And the strange new world of nonmonotonic reasoning was beginning to be explored (Bobrow, 1980).

Despite the beginnings of more widespread formalization and some basic standard apparatus for frame systems, the picture painted by the community-wide survey in (Brachman & Smith, 1980) was still one of *ad hoc* methodology and heterogeneity. Most KR groups were building their own systems, which were used only by themselves, and there was still significant disagreement on many fundamental issues.

Important Developments of the 1980's

Since 1980 many things have changed. The last decade has been impressively productive, there is widespread agreement on many issues, and the methodology in the field is much more uniform. I will outline what appear to be the most important developments since 1980, although any such catalogue is an oversimplification along many dimensions.

Overall Trends

There are some general trends that have characterized KR research over the last 10 years that are almost unanimously perceived as significant (although not uniformly as positive).

Technical sophistication. Over the last few years, KR papers have become increasingly technical. By and large this is good—it seems to represent a decline in the kind of loose meta-discussion popular in the 1970's and an increase in interest in getting some real work done. On the other hand, this increase in technical detail has made KR work much harder to comprehend for those not already involved, and has helped draw KR away from the other areas of AI with which it had traditionally been allied (e.g., natural language).

[5] See (Brachman & Smith, 1980) for an extensive but non-homogenized catalogue of the field as of the late '70's; see also (Findler, 1979) and (McDermott, 1978).

[6] See (Hayes, 1977) for a summary of the arguments. Hayes states that the proceduralists' arguments were conclusive, although he goes on to point out serious misunderstandings of the foundations of KR in that debate.

Theory. KR work in the last ten years has also increasingly focused on theory. Concomitantly, KR discussions in the literature have been approached with substantially more rigor than they were in the past. Some impetus for this came from several prize-winning papers at AAAI conferences in the mid-1980's (when interest in AI was at its peak) and a groundswell of effort on formalisms for nonmonotonic reasoning. More theorems and proofs than ever have appeared in recent KR papers and the body of mathematics in support of KR has grown dramatically. A formal semantics is now an obligatory accompaniment of the description of a novel KR system.

The tremendous upsurge in KR theory has seemingly come at the expense of experimentation in the field, which has been discouraged by the actions of various program committees. This is not to say that the "experimental" side of KR (mainly associated with work on computer programs) could not use an injection of rigor itself. But by most accounts, the amount of empirical work in the area has fallen off, or has gone underground.

To the good, the injection of theory and rigor allows us to be able to determine if programs really work and has given us KR systems with precise formal semantics—a clear improvement over work typical of the pre-1980 era. But the pendulum may have swung too far, inadvertently causing a rift between the formalists and those concerned with applications, and causing less and less of the KR literature to have any impact on the rest of AI and on practice.

Declarative representations. The community has moved away from "procedural representations" to a radically declarative worldview. KR work has increasingly focused on what our representations *say* (e.g., about the world) and less on how to control procedures that process them. The ultimate embodiment of this view is standard first-order logic with Tarskian semantics. A decade ago, debate raged on the utility of logic and declarative semantics, and on whether representations had to be "programmable"; at that time, for example, "procedural attachment" was *de rigueur* in frame systems. While no doubt there are still those who wish to resurrect the issue, the field as a whole seems to have warmly embraced classical logic with standard model-theoretic semantics. This and related issues were discussed at length in a special journal issue (Levesque, 1987) addressing McDermott's critique of "pure reason," but even McDermott, while unsure of the prospects for success of use-independent, purely declarative knowledge bases, could not come up with a very concrete alternative. Major contributions to the declarative view included Levesque's functional approach to knowledge representation (Levesque, 1981) and Newell's "knowledge level" proposal (Newell, 1981).

Reasoning. Almost paradoxically, there has been increased focus in the KR community on general types of reasoning. Attention has shifted from the preoccupation with pure language design of the '70's to the arena of different types of inference (e.g., temporal, "abductive," "case-based"). This is a good sign, in that the older formalisms could rarely be evaluated because it was never clear what they meant or what inferences they sanctioned. This however, has led to some difficulty, in that conference calls for papers continue to divide KR from its *sine qua non*, reasoning (e.g., "Commonsense Reasoning," "Automated Reasoning," and KR are considered distinct topics). Ten years ago, this split might have been appropriate, but now such distinctions are artificial and detrimental. It has become hard to tell where most papers belong and how to judge them.

Concentration. Another trend of note is a terrific concentration—almost an "implosion"—of work in a small set of areas, most notably nonmonotonic reasoning and qualitative reasoning about physical systems. A look at recent conference proceedings reveals how large a number of people have flocked to a very small set of issues.

Technical Developments

Within the context of the above global trends, there have been numerous notable technical developments in KR over the last decade. Here I briefly gloss over some of the more obvious happenings to give a flavor of what caught the fancy of the KR community.

Nonmonotonic reasoning. Easily the most noticeable KR area at recent conferences has been that of "nonmonotonic reasoning" (NMR), wherein formal approaches are proposed to handle the fact that a great deal of reasoning must be based on assumptions that may be ultimately shown to be false. Since information learned at some future time may falsify a default assumption, many commonsense reasoning situations are inherently nonmonotonic.[7]

The last ten years has seen the introduction of numerous formalisms for NMR, including circumscription, Default Logic, autoepistemic logics, conditional logics, and many variants of inheritance systems. Some of these are semantic, or consistency-based systems (e.g., they depend on a certain default being consistent with an entire KB) and some are syntactic (e.g., they depend on paths through a graph). A key insight has been the use of "minimal models" as the semantic basis of many nonmonotonic systems. Among the key developments in the '80's was work by Reiter and students showing that some flaws in inheritance mechanisms could be elucidated by formalizing the networks in terms of Default Logic, and that the commonly used "shortest-path heuristic" was inadequate.

The nonmonotonic world has concentrated on a small number of canonical problems, such as deter-

[7]For a more comprehensive introduction to this area, see (Etherington, 1988), (Reiter, 1987), and (Ginsberg, 1987).

mining if a given bird can fly from the statement of a default like "typically, birds fly." This is harder than it may look, if you want to take seriously the possibility of birds not flying. Once the basic problem is solved, other more convoluted ones arise, such as the multiple-extension, "Nixon Diamond" problem (in this case the typical NMR system will produce two possible answers to a default inheritance problem). Another key development (although the importance of the problem itself is arguable) was the "Yale Shooting Problem" (Hanks & McDermott, 1986), which illustrated how circumscription could yield counterintuitive results on certain problems involving the projection of events over time. This work caused quite a stir even before it appeared in print, and has provoked work on many solutions.

The Yale Shooting Problem and other canonical NMR problems involve a very small number of axioms to describe their entire world. These may not be fair problems because the knowledge involved is so skeletal. It seems unrealistic to expect a reasoner to conclude intuitively plausible answers in the absence of potentially critical information. By and large, NMR techniques have yet to be tested on significant, "real-world"-sized problems.

Qualitative physics. As mentioned, in the '80's there was tremendous growth in interest in qualitative reasoning, almost exclusively about continuous physical systems (thus it is usually called "qualitative physics" (QP)) (Bobrow, 1984; Weld & de Kleer, 1990). The work has its roots in work on engineering problem-solving at MIT in the 1970's, Hayes's work on Naive Physics, and in early work on how devices work by Rieger and Grinberg. As with other fields, QP does not have a single coherent view, and people differ on the key goals. But generally, QP is about how things work, including (1) what are appropriate representational primitives for the salient features of devices and their behavior? (2) given a physical artifact and an initial situation, how do we construct a description of how that artifact works? and (3) how do we use this description to perform interesting tasks like diagnosis and design? Most agree that explanations of how things work have qualitative, temporal, and causal components.

The first half of the '80's was spent roughing out the representations of these explanations, and proposing techniques for generating them. The greatest advances were on the qualitative component, although work proceeded on the others. The main issue is, what are good qualitative representations for the values of quantities and relationships between quantities? Originally, the sign of the derivative of a quantity was proposed as an important qualitative representation; this was expanded to include ordinal relations ($<$, $=$, $>$) of "landmark values," and eventually to have a device's "state space" divided into open regions of interest separated by boundaries, where "regions of interest" are defined by the task. Representations for relationships between values were explored, among them, "confluences"—equations on the signs of quantities.

In the mid-'80's, work in QP took off, in part spurred by a special issue of *Artificial Intelligence* (Bobrow, 1984). At the center of the more recent work has been representations for relationships between values, including more powerful qualitative algebras, and more direct ties to the standard frameworks of mathematics and physics. There is also, of course, a tremendous amount of work on reasoning with these and other representations, which we do not have space to cover here.

The notion of a truth maintenance system (TMS) (McAllester, 1990) grew up within the community interested in reasoning about physical devices, although the ideas have now spread far and wide. The ideas go back to early work at MIT. Interestingly, this work ties directly back into work on nonmonotonic reasoning. Work on TMS's grew tremendously in the '80's.

While NMR and QP shared the KR limelight in the '80's, they were not the only technical activities. A number of other topics with keen interest emerged:

The revival of probabilities. In the 1970's, it was virtual heresy to talk about numbers, since that was taken to mean that something was being swept under the rug (e.g., causal relationships disappeared in numerical approaches, and even in the medical AI community, many turned away from probabilities). While Bayesian/decision-theoretic approaches were attractive, since they allowed one to maximize expected value even in cases that were not statistically significant, they were problematic because they either required you to assume that everything was conditionally independent or that everything was dependent. In the '80's work on Bayes networks (Pearl, 1988) yielded representations that could express partial dependence and partial independence. This allows one to tackle large-scale decision problems from a formal probabilistic perspective, and it should have some important practical implications. Other important work integrating probabilistic and deductive reasoning is just beginning to unfold, and could have a major effect on KR work on realistic problems in the next few years.

Hybrid reasoning systems. By the mid-'80's, combining multiple types of representation was popular. In a division of representational labor, specialized subsystems stitched together might provide the power to handle realistic domains without forcing a single uniform, too-powerful logic. Hybrids of various sorts were developed, including several marrying logic and frames. Sorted logics grew in popularity, and commercial expert system shells generally offered several loosely integrated types of representation. One important criterion for separation of hybrid components (Brachman & Levesque, 1982) distinguished between terminology (knowledge about the meanings of terms, independent of the existence of any objects exemplifying those terms) and assertion (knowledge of contingent facts). A large family of terminological and hybrid systems developed inspired by KRYPTON (Brachman *et al.*, 1983) (these ultimately attribute their roots to work done on KL-ONE (Brachman & Schmolze, 1985) in the late '70's and early '80's). Novel work was also done on taxonomic syntax and reasoning.

Complexity of reasoning. An influential piece of work that grew out of hybrid representation involved proofs of the computational complexity of the term subsumption inference (Brachman & Levesque, 1984). The results were surprising: small syntactic changes in a representation language can lead to dramatic changes in inference complexity ("computational cliff" was a term used to describe the transition). There soon followed a large number of analyses of the complexity of reasoning with various related term-subsumption languages. In the latter half of the decade, it became common to see complexity results for reasoning in various types of systems, ranging from default inheritance schemes to abduction. Besides leading to systems whose complexity characteristics were understood, this led to better appreciation of how hard it is to avoid potentially intractable inferences.

Case-based reasoning. Another important thread in the '80's has been the attempt to reason based on catalogues of past experience, appropriately indexed. By drawing analogies to the current situation, a case-based reasoner may be able to reuse or revise a previously stored solution. This work has looked at memory structures for case-based reasoning in general and in legal argumentation, medical diagnosis, etc.

Abductive reasoning. Throughout the decade, in contexts ranging from story understanding to circuit diagnosis, attention was paid to foundations of diagnostic reasoning, usually called "abductive" inference. Abduction is reasoning to the best explanation—a nonmonotonic inference. Recently, general algorithms for abduction have been studied, and the complexity of abductive inferences have been catalogued. In some cases, diagnostic reasoning "from first principles" (sometimes called "model-based diagnosis") has been achieved—an important advance over shallower expert system diagnostic methods. Interest in abduction has been around since the early '70's, but was dormant until recently. Interest in diagnosis is growing rapidly.

Vivid reasoning. In 1985, Levesque (Levesque, 1986) introduced a novel approach to reasoning that attempts to deal with the complexity of reasoning by allowing the relaxation of correctness and completeness in some cases. The basic idea is to use past experience and other default knowledge to reduce general first-order knowledge to roughly database—or "vivid"—form. Reasoning in the vivid KB then reduces to fast retrieval, although some information must be lost and some errors introduced in the translation. This idea has led to several interesting technical innovations.

Large KB's. In the '80's we began to see the rise of projects developing very large knowledge bases. For example, Lenat and colleagues initiated the CYC project (Lenat & Guha, 1990)—a bold attempt to encode millions of fragments of "consensus reality" in an encyclopedic knowledge base. This project has begun to raise a host of issues previously unaddressed, simply because of its magnitude and its need to concern itself with "ontology." More generally, with the contemplation of significant investment of time and energy into single KB's, concern has begun to develop about the reusability of KB's, knowledge base management issues, general issues of ontology, and standards for representation languages. A recent workshop explored the possibility of developing an interlingua that could be used to share KB's from one project to another, even if the projects used different KR languages.

Other. There were a number of other topics pursued rather vigorously by the KR community, among them temporal reasoning, "model-based" reasoning, reasoning about mental systems (including intensionality, goals and commitments, explicit and implicit belief, and combining evidence), and continued and expanding work on a few network representation systems (e.g., Conceptual Graphs, SNePS).

Other Developments

Beyond its own technical progress, KR benefited from the commercialization of AI in the 1980's. It began to have an impact on the "real world" via expert system shells sold to the public and used in commercial applications (even *Byte Magazine* had a special issue on KR). While the KR technology that supported most commercial expert systems work was somewhat simple and had been developed long before, some of the larger shells included multiple representational components (typically rules, frames, and some logical representation). With rule-based programming becoming an acceptable alternative (not to mention the widespread use of PROLOG, which bears some relationship to KR), KR made its way into the mainstream of technology.

Another intriguing development of the 1980's was the beginning of exploration of connections to other fields, including decision theory, control theory, economics, etc. (e.g., (Doyle, 1990)).

And finally, in 1989, the KR'89 conference (Brachman *et al.*, 1989) brought together the KR community.

Nagging Doubts

One of the most important developments for KR has been the recent resurrection of interest in connectionist architectures. There are many brands of connectionism, some of them quite compatible with the KR view; one key strain advocates non-symbolic computation. The statistical nature of some connectionist systems makes them less fragile than symbolic KR systems in the face of noise, and since ultimately much of the information to be obtained by AI machines will be noisy, this seems to indicate that connectionist systems will eventually take over the role now being played by traditional KR systems. Arguments along these lines have been made, but the jury is still out. Connectionist systems have been found useful for low-level, more perceptual tasks (like handwritten character recognition), but are still a long way from being able to produce a plan or an explanation. Ultimately, we will probably see hybrid systems with both connectionist and symbolic components, but for now, there is widespread debate about what connectionist systems will eventually be able to do, and what kind of threat they are to more traditional approaches.

In a somewhat related vein, some researchers have claimed that many types of intelligent-seeming behavior can be accomplished without the use of explicit knowledge, but rather with compiled-in structures that allow a system to react directly to its environment (to "lean on the world" for parts of its knowledge, rather than store it explicitly internally). The "reactive" and "situated" (Rosenschein, 1990) (where the fact that the system is embedded in its world is taken as primary) approaches are concerned with real-time performance, and in some cases appear able to overcome computational bottlenecks of approaches that use theorem-proving. Again, there is an issue of how much of intelligent behavior is achievable with purely reactive systems, although there are arguments to be made about how such architectures more closely resemble (parts of) natural systems.

Even within the hard-core KR community, there were doubts about traditional ways of doing things. The "logicist" approach—strict declarativism, with knowledge represented independent of use—was taken to task by McDermott (McDermott, 1987). This sparked continued debate (Levesque, 1987) over whether it is desirable or productive to build large knowledge bases without direct attention to the intended application of the knowledge (not to mention continued argument on the role of logic in KR).

All of this provides some healthy skepticism for a field that could easily get set in its ways. It is important to remember that there are situations where symbolic representation is either implausible or inadequate. In fact, while some KR people might think of connectionism as a radical hypothesis, when one looks at the natural world, it becomes apparent that the symbol-manipulation view of intelligence—the well-spring of KR—is the more radical view.

Some "Non-Happenings"

It is interesting to reflect on events that might have been predicted for the '80's, but which never took place. These items still hold the promise that they did a while ago, and in most cases interest seems to be returning. But they were surprising by their absence.

AI–DB integration. In the mid-'80's, the prospect of "expert databases," and the commonplace integration of KR and DB technology excited great interest. Work on deductive databases held the promise of synergy with mainstream KR. So far, this has not panned out, perhaps because the basic concerns of practical database management and KR are quite different.

"Prototypes." Minsky's original frames paper and work on KRL promised that insights from cognitive science on prototypes and basic categories would have a major influence on AI reasoning systems. Unfortunately, this and other aspects of the frames paper seem to have gotten lost along the way.

Natural language semantics. Given that KR was principally driven by natural language concerns right up to the beginning of the decade, one would have expected substantial progress to have been made in the '80's on KR support for NL semantics. This seems not to have been the case.

Diagrammatic representations. In 1980, Funt (Funt, 1980) presented an innovative but under-appreciated representation for visual information. Collisions of objects could be detected by direct calculation on an analogue representation.

"What Computers Can't Do." Throughout the history of KR, there has been debate over whether thinking can be achieved by a mechanized process that manipulates symbols. The revised version of Dreyfus' book and Searle's account of the "Chinese Room" seemed to portend great difficulty for KR practitioners. As it turns out, regardless of the ultimate cogency of the arguments against formal AI, work in KR has proceeded without heed.

Meta-reasoning. Early in the '80's there was much hope that many hard problems could be solved by "going meta." So far, meta-reasoning has not turned out to be a panacea.

The Future of Knowledge Representation

Despite the telegraphic nature of the above comments, it should be clear that in the past ten years KR has seen some significant changes. Work is substantially more formal and rigorous than it was prior to 1980, fewer implemented systems are being discussed, and a small number of issues have absorbed great attention. The field has moved to center stage in AI, owing to several factors, including some "best papers," journal special issues, Computers and Thought lectures, eye-catching problems and projects, widespread use of expert system shells, and a dedicated international conference.

But there are some lurking worries. Many feel that the emphasis on formal logic has gone too far, and that important experimental work is being squeezed out by purely theoretical concerns. Much of the work being done bears little regard for realistic problems. The pipeline of interesting problems to analyze (usually generated by attempts to build systems) seems to be filling too slowly. As KR focuses more on self-generated technical problems, it seems to move farther from the rest of AI that it was originally intended to support. And connectionism and other non-symbolic approaches seem to pose a threat to the very existence of the KR enterprise.

What does this portend for the next ten years? Having given some thought to major developments of the past, it is time for us to get back to the future of knowledge representation. Here are some potential highlights.

Some Likely Scenarios

Logic and rigor. The emphasis on rigor in KR will probably continue. Despite some disgruntlement in various circles it is also likely that the logicist program will continue to dominate, although we might see a procedural backlash of some sort, and serious discussion of the role that intended use plays in the form of knowledge. Some believe that there will be movement away from classical model-theoretic semantics, and perhaps more interest in relevance logics and intuitionistic logics. There will continue to be some concern about computational complexity—this is healthy, although the meaning of such results needs to be clarified (see below), and finer-grained analyses are needed.

KR in general will show an increasing frustration with seemingly irrelevant mathematics and theorems not clearly motivated by important problems. While the mathematical foundations of the field are critical, the talented community interested in such work may seek a more comfortable home (e.g., a specialized conference); movement of this sort is already underway.

Nonmonotonic reasoning. Intuitions about the meaning of nonmonotonicity and its different incarnations will probably become better identified. We will see a wave of formal systems based on arguments for and against a given conclusion; ultimately, we should see work relating such systems back to the consistency-based and syntactic systems. In general, if we are lucky, the field will realize that it is the skeletal, impoverished-knowledge cases on which different views clash, and concern itself with more conservative approaches; that is, if most NM systems agree on the common cases, then there should be less argument about which is superior and which is flawed, and more on how to solve the really fundamental problems. All in all, this could result in fewer new approaches and more utility from the ones we have, and a revived concern for solving real problems with realistically complex KB's (this should ultimately make the problem easier, not harder). Again, if we are lucky, we should start seeing analyses of the nonmonotonic aspects of real-world problems, and the proposal of limited and conservative mechanisms that actually solve them.

Unified reasoners. In the near future, we can expect to see more theories that unify different types of reasoning. For example, we are beginning to see logics that incorporate both deduction and abduction (diagnosis and explanation are integrated into a deductive framework). This trend is likely to continue, with, for example, induction added to the arsenal of reasoning strategies in a unified system. Other efforts (e.g., (Bacchus, 1990; Halpern, 1989)) have begun to unify deductive, nonmonotonic, and probabilistic reasoning in a coherent and smooth way. Ultimately, reasoning about action and time (planning) and the kind of reasoning underlying learning will also find their way into what might ultimately become a grand unified theory of reasoning.

Probability and statistics. We can expect work on statistical and probability-based reasoning to become more closely associated with mainstream KR. We will soon see more accounts of probabilistic and fuzzy reasoning that are compatible with standard approaches. These will help meet the challenges to deductive reasoning currently posed by the need to handle noisy data, frequency information (e.g., in learning), and "fuzzy" concepts. In ten years, many approaches will include both a standard deductive/categorical mechanism and one for dealing with limited observations of regularities in the world (in fact, the deductive mechanism may be the minor component, used only in sticky, "puzzle-mode" situations). We are seeing signs of this in NL, wherein statistics about co-occurrence of words can provide valuable disambiguating information for more traditional categorical approaches to parsing, and it is likely that the influence of probability will have a similar effect on KR and reasoning systems. This kind of approach will be critical in planning, for example, where the view of a plan as a theorem to be proved is too restrictive to be realistic.

Natural language. There has been a deep connection between KR and NL as long as those areas have been studied, and the connection is still there, even though KR research has focused less on NL-specific issues (with some recent movement in the NL community towards statistical text-based analyses, they have drifted even farther apart). Some predict, however, that we will see a strong move back to support NL semantics and pragmatics by some parts of the KR community. Some recent work (Schubert & Hwang, 1989) illustrates the potential direction of this work. Older work by Martin (Martin, 1979) will probably be reexamined and found surprisingly innovative and useful. The issue of indexicality will continue to be addressed, although it is not clear that representations need to be indexical the way that language is.

Ontology. Issues of ontology will be among the most important and most talked about in the next few years. How to build the "upper model"—the topmost levels of a large hierarchy of commonsense knowledge, how to integrate parts created by different people, and how to control revisions will be important considerations; the automatic generation of ontologies will probably also become a much-discussed topic. Along related lines, "prototypes," basic categories, and other psychological phenomena having to do with categorization of the world will probably make a comeback.

Large KB's. Very large knowledge-based systems will soon be commonly upon us. With this, issues that have occupied the database world will come to concern KR developers, although perhaps complicated in interesting ways by the logical interpretation of KR languages. Among such concerns, we might see sharing of KB's, persistent object stores, dealing with outdated or suspect information, "drift" of terminology, and infrastructure issues such as ownership and commercial value of represented information. Incremental revision will be of paramount importance, since KB's will exist over longer periods of time. Knowledge acquisition will of necessity move closer to mainstream KR—with very large knowledge bases, automatic and semi-automatic acquisition will become the *sine qua non* of KB's. Memory organization issues will reappear within KR (some of the earlier work of Schank, *et al.*, will possibly be rediscovered or reinvented). Finally, much will be learned from trying to build CYC, although serious obstacles (e.g., difficulty of timely inference in a large enough KB, reconciliation of pieces built by different authors, and general skepticism of the utility of a KB built without a particular use in mind) may prevent it from being anything other than an exciting first experiment.

KR to the people. It is probable that by the millenium "knowledge systems" will be a common commercial concept. This has important implications for the future of KR. Among other things, KR components will increasingly find themselves in the hands of non-experts, raising a novel set of issues (see below). The issue of KR standards will grow somewhat in importance, with immediate emphasis on intertranslatability between various languages; no doubt subcommunities will also work on true standard KR languages for certain classes of problems. I suspect that, because of the diversity of approaches in the field and current lack of incentives for standardizing, the results of this standards work will be less than ideal for the foreseeable future. However, with increasing governmental and commercial investment in knowledge bases, it really does need to be addressed.

No HAL. It is hard to ignore the fact that in roughly another decade we will be confronting the year 2001. Sad to say, given the progress to date, we can comfortably predict that there will be no HAL-9000 available (or even anything close).

Some Open Research Problems

I now briefly cover some issues in need of research, some broad and some technical. Some of these may have work underway already, so the distinction between the above predictions and this wish list is not absolute.

Expressiveness vs. tractability. The general issue of computational efficiency needs more work. What is the meaning of the complexity results produced so far? Worst-case results seem too coarse-grained to be of use in designing real systems, although without understanding how often such cases arise, or what a normal case might be, we still need to pay careful attention to such results. Perhaps decision-theoretic methods could be used as a way of dealing with intractable problems.

An irony of work on NMR is that, while the easy adoption and retraction of assumptions is most useful for speeding up natural everyday reasoning, most current NMR proposals drastically compound the already difficult problem of deductive reasoning. We urgently need to determine how NMR can be used to make commonsense inference faster, not slower. Unless progress can be made on this front, then most of NMR will probably end up as an interesting mathematical dinosaur. We need ways of doing quick, perhaps inaccurate (but reasonable) reasoning, and to understand when and how to fall back on more reliable, but slower methods.

Incomplete reasoners. As we build more expressive KR systems, we are virtually guaranteed that they will be incomplete. What is the most useful way to build an incomplete reasoner? Are there ways of describing such systems so that users will understand exactly what to make of the results returned and how much to trust the system at any point? Can we build systems that yield results any time we need them (i.e., the best guess at the time) and whose results improve as we let them run longer?

KR services. It is possible that the idea of a "general-purpose KR system" that has pervaded the field for many years is meaningless. It may not be possible to rationalize the needs of all applications in one system, and simply opting for maximal expressive power may not be the best strategy in all situations. When can the needs of a KR service be constrained? Are there different reasonable and natural levels of service that can be provided? (I would like to see these levels characterized in a "knowledge level" way, rather than simply saying what procedures are invoked at what level.) Can one system provide many different levels at the user's option? Can we characterize the cost of various services so that a user can take it into account in deciding whether to invoke an inference mechanism? In general, the roles that KR components of a knowledge-based system can play need to be articulated, and the services that the KR system is supposed to provide should be clarified. In the past we have emphasized the need for a KR component to have "predictable" behavior. If "predictability" does not mean "completeness," what exactly does it mean?

KB management. With the advent of very large, realistic knowledge bases, we will need to address the problem of inconsistency: how can a knowledge-based system live with globally inconsistent, but locally reasonable knowledge? The problem of limiting the scope of an inference procedure will also be important. This is not a new issue (how to find the relevant things to think about in a vast sea of undifferentiated facts), but it is likely to have practical consequences in the next ten years. Knowledge base management issues will grow in importance (including low-level ones like version control). More attention will need to be paid to a higher level of knowledge organization, "above" the knowledge level: what are good principles for organizing knowledge of various sorts? And, since large KB's will invariably be built by multiple sources over much time, we will also need practical but well-founded theories of belief revision.

Usability of KR systems. As mentioned, KR systems will increasingly find their way into the hands of "non-professionals." How do we ease this market entry? What aspects of KR need to be emphasized and which de-emphasized in order to allow real users to deal with the knowledge and not get bogged down in system details or complex KR issues that seem irrelevant to them? Experience from the commercial world of expert system shells should be heeded here. Issues of presentation of knowledge, browsing, and querying will be critical, as will the need to explain the system's reasoning in terms comprehensible to normal humans. KR systems in general are likely to become much more complicated, especially as more types of reasoning become better understood and are routinely incorporated; the inner workings will thus become incomprehensible even to experts and these same issues will be doubly important.

Other. There are numerous other research problems on which work is needed; here are a few practical ones: integration of frame classification systems and respectable theories of defaults and inheritance; precise theories of approximate categorization, ultimately integrated with more deductive classification schemes; extensible representation schemes that allow expert users to add important constructs, without having to rebuild the system from scratch; the efficiency and usability of reason maintenance systems (critical as we scale up).

Deeper issues. Finally, there are of course deeper and more general issues. For example, there seems to be some strong general sentiment for the need to take into account the use to which knowledge is to be put before designing a representation or making any claims about it. Can we have realistically useful KB's that are designed in absence of specific intended applications?

Why are we so convinced that knowledge actually has to be represented at all? What kinds of activities are infeasible without general reasoning abilities? Are there principles for making something explicit or declarative? Current work on reactive and related systems will hopefully produce vital data on the limits of non-declarative systems (if there are any).

In fact, in the long run, we might consider whether the idea of a separable KR component really makes sense at all. How could the mind have such "loose coupling" between its "knowledge base" and the rest of its capabilities? Recent work in databases, as well as situated automata, reactive systems, and neural nets indicate that a tightly integrated reasoning system may be more realistic than one with the typical bipartite knowledge-based architecture.

Some General Recommendations

One of the imminent dangers of work in KR is the risk of losing touch with the rest of AI, despite the fact that the *raison d'être* of KR is to support it. As we delve into more technically sophisticated details, we seem to leave our "customers" farther and farther behind. While we need to address our problems with technical depth and rigor, we also need to avoid the syndrome of "epicycles" and be careful about losing the forest of commonsense reasoning for the trees of default inheritance. It would not hurt at this point to go back and spend time thinking about the relation of KR to natural language, for example—after all, that was in part responsible for the birth of the field in the first place. This is not to say that KR should return to its primordial ooze, but only that we must refocus on real and important problems. This is especially true in areas like NMR and QP: what real-world problems are better off now than they were ten years ago?

We also need to re-encourage experimentation, although we can be very careful about what constitutes an acceptable experiment. In a way, we need to get back to the "pioneering" spirit of the 1960's and '70's, but armed with the insight and mathematical arsenal of the '80's. Conference program committees in KR should look for more good ideas and fewer mathematical journal-style papers. Our conferences have taken on too much of the flavor of journal-readings and have lost some of the excitement of actually *conferring* to argue about new and provocative ideas.

In order for this to be realistic, systems-oriented people should give some serious thought to what constitutes a result in the experimental side of the field. We need more consideration by this community as a whole as to what its goals are, and what the important issues of KR system design and implementation are. The development of some metrics for measuring quality, scope, etc., would be especially welcome.

We should try to eliminate the insidious split that has developed in calls for papers—since commonsense and other forms of reasoning are the very reason that KR exists, it is problematic to list them as completely separate areas. It is hard under the current scheme to know to which area to send papers. We must remember that KR stands for knowledge representation *and reasoning*, and structure our topics accordingly.

We should also continue developing relationships with mainstream computer science and other disciplines that are related to our enterprise, including control theory, decision theory, statistics, OR, economics, etc. We must remain open-minded about input from these other disciplines. Recent work has indicated that the payoff from merging AI ideas with those of more traditional disciplines could be great.

All in all, it is clear that KR has been thriving these past ten years: it has developed a substantial body of technical machinery and has moved into the "real world." The next decade promises to be as intellectually rich as the last, and if the right attitude is taken and the right problems addressed, it should bring us much closer to having reliable, fast, and reasonable reasoners.

References

Bacchus, F. 1990. *Representing and Reasoning with Probabilistic Knowledge.* The MIT Press, Cambridge, Massachusetts.

Bobrow, D. G., ed. 1980. Special Issue on Non-Monotonic Logic. *Artificial Intelligence,* 13(1–2).

Bobrow, D. G., ed. 1984. Special Volume on Qualitative Reasoning about Physical Systems. *Artificial Intelligence,* 24(1–3).

Brachman, R. J., and Levesque, H. J. 1982. Competence in knowledge representation. In *Proc. AAAI-82,* pages 189–192, Pittsburgh, PA.

Brachman, R. J., and Levesque, H. J. 1984. The tractability of subsumption in frame-based description languages. In *Proc. AAAI-84,* pages 34–37, Austin, TX.

Brachman, R. J., and Levesque, H. J., eds. 1985. *Readings in Knowledge Representation.* Morgan Kaufmann, San Mateo, California.

Brachman, R. J., and Schmolze, J. G. 1985. An overview of the KL-ONE knowledge representation system. *Cognitive Science,* 9(2):171–216.

Brachman, R. J., and Smith, B. C., eds. 1980. Special issue on knowledge representation. *SIGART Newsletter,* No. 70.

Brachman, R. J., Fikes, R. E., and Levesque, H. J. 1983. KRYPTON: A functional approach to knowledge representation. *IEEE Computer,* 16(10):67–73.

Brachman, R. J., Levesque, H. J., and Reiter, R., eds. 1989. *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning.* Morgan Kaufmann, San Mateo, CA.

Doyle, J. 1990. Rationality and its roles in reasoning. In *Proc. AAAI-90,* Boston, MA.

Etherington, D. W. 1988. *Reasoning with Incomplete Information.* Pitman, London.

Findler, N. V., ed. 1979. *Associative Networks: Representation and Use of Knowledge by Computers.* Academic Press, New York.

Funt, B. V. 1980. Problem-solving with diagrammatic representations. *Artificial Intelligence,* 13(3):201–230. Also in (Brachman & Levesque, 1985).

Ginsberg, M. L., ed. 1987. *Readings in Nonmonotonic Reasoning.* Morgan Kaufmann, San Mateo, CA.

Halpern, J. Y. 1989. An analysis of first-order logics of probability. In *Proc. IJCAI-89,* pages 1375–1381, Detroit, MI.

Hanks, S., and McDermott, D. 1986. Default reasoning, nonmonotonic logics, and the frame problem. In *Proc. AAAI-86,* pages 328–333, Philadelphia, PA.

Hayes, P. J. 1977. In defence of logic. In *Proc. IJCAI-77,* pages 559–565, Cambridge, MA.

Lenat, D. B., and Guha, R. V. 1990. *Building Large Knowledge-Based Systems.* Addison-Wesley, Reading, MA.

Levesque, H. J. 1981. The interaction with incomplete knowledge bases: A formal treatment. In *Proc. IJCAI-81,* pages 240–245, Vancouver, British Columbia.

Levesque, H. J. 1986. Making believers out of computers. *Artificial Intelligence,* 30(1):81–108.

Levesque, H. J., ed. 1987. Taking Issue/Forum: A Critique of Pure Reason. *Computational Intelligence,* 3(3).

Martin, W. A. 1979. Descriptions and the specialization of concepts. In P. H. Winston and R. H. Brown, eds., *Artificial Intelligence: An MIT Perspective, Vol. 1,* pages 377–419. MIT Press, Cambridge, MA.

McAllester, D. 1990. Truth maintenance systems. In *Proc. AAAI-90,* Boston, MA.

McDermott, D. V. 1978. The last survey of representation of knowledge. In *Proc. AISB/GI 1978,* pages 206–221.

McDermott, D. V. 1987. A critique of pure reason. In (Levesque, 1987), pages 151–160.

Minsky, M. 1975. A framework for representing knowledge. In P. H. Winston, ed., *The Psychology of Computer Vision,* pages 211–277. McGraw-Hill, New York. Also in (Brachman & Levesque, 1985).

Newell, A. 1981. The knowledge level. *AI Magazine,* 2(2):1–20.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, CA.

Reiter, R. 1987. Nonmonotonic reasoning. In *Annual Reviews of Computer Science, Volume 2,* pages 147–186. Annual Reviews Inc., Palo Alto, CA.

Rosenschein, S. J. 1990. Reasoning and acting in real time. In *Proc. AAAI-90,* Boston, MA.

Schubert, L. K., and Hwang, C. H. 1989. An episodic knowledge representation for narrative texts. In (Brachman *et al.,* 1989), pages 444–458.

Weld, D. S., and de Kleer, J., eds. 1990. *Readings in Qualitative Reasoning about Physical Systems.* Morgan Kaufmann, San Mateo, CA.

Woods, W. A. 1975. What's in a link: Foundations for semantic networks. In D. G. Bobrow and A. M. Collins, eds., *Representation and Understanding: Studies in Cognitive Science,* pages 35–82. Academic Press, New York. Also in (Brachman & Levesque, 1985).