Logic and Artificial Intelligence

Richmond H. Thomason
Department of Philosophy
University of Michigan
Ann Arbor, MI 48109–2210
USA

rich@thomason.org
www.rthomaso.eecs.umich.edu

October 23, 2002

This is a working draft.
Do not cite.
Comments welcome.

Version date: October 23, 2002

1. Logic and Artificial Intelligence

Artificial Intelligence (which I'll refer to hereafter by its nickname, "AI") is the subfield of Computer Science devoted to developing programs that enable computers to display behavior that can (broadly) be characterized as intelligent. Most research in AI is devoted to fairly narrow applications, such as planning or speech-to-speech translation in limited, well defined task domains. But substantial interest remains in the long-range goal of building generally intelligent, autonomous agents.[1]

Throughout its relatively short history, AI has been heavily influenced by logical ideas. AI exhibits a rather eclectic assortment of theories and research methodologies; the value and relative importance of logical formalisms is questioned by some leading practitioners, and has been debated in the literature from time to time.[2] But most members of the AI community would agree that logic has an important role to play in at least some central areas of AI research, and an influential minority considers logic to be the most important factor in developing strategic, fundamental advances.

1.1. Guide to this Chapter

I imagine that the audience for this chapter will consist primarily of logicians and historians of logic who have little or no familiarity with AI. In writing this chapter, I have tried to provide an overview of the issues that arise when logic is used in helping to understand problems in intelligent reasoning and to guide the design of mechanized reasoning systems. Logic in AI is a large and rapidly growing field—I could not hope to achieve anything like complete coverage. In Sections 3 and 4 I have tried to provide an overview with some historical and technical details concerning nonmonotonic logic and reasoning about action and change, a topic that is not only central in AI but that should be of considerable interest to philosophers. The remaining sections provide brief and inadequate sketches of selected topics, with references to the primary literature.

While this chapter was being written, [Minker, 2000b] appeared. This book is a comprehensive, up-to-date collection of survey papers and original contributions to the field of logic-based AI, with extensive references to the literature and with an introduction (to the book and to the field), [Minker, 2000a]. I highly recommend this volume as a beginning point for any readers who wish to pursue this topic further.

1.2. The Role of Logic in Artificial Intelligence

Theoretical computer science developed out of logic, the theory of computation (if this is to be considered a different subject from logic), and some related areas of mathematics.[3] So theoretically minded computer scientists are well informed about logic even when they aren't logicians. Computer scientists in general are familiar with the idea that logic provides techniques for analyzing the inferential properties of languages, and with the distinction between a high-level logical analysis of a reasoning problem and its implementations. Logic,

[1] See, for instance, [Nilsson, 1995].

[2] For two debates, see Volume 3, Number 3 of *Computational Intelligence*, devoted to [McDermott, 1987], and the later exchange [Nilsson, 1991, Birnbaum, 1991].

[3] For some of the historical background, see [Davis, 1988].

for instance, can provide a specification for a programming language by characterizing a mapping from programs to the computations that they license. A compiler that implements the language can be incomplete, or even unsound, as long as in some sense it approximates the logical specification. This makes it possible for the involvement of logic in AI applications to vary from relatively weak uses in which the logic informs the implementation process with analytic insights, to strong uses in which the implementation algorithm can be shown to be sound and complete. In some cases, a working system is inspired by ideas from logic, but acquires features that at first seem logically problematic but can later be explained by developing new ideas in logical theory. This sort of thing has happened, for instance, in logic programming.

In particular, logical theories in AI are independent from implementations. They can be used to provide insights into the reasoning problem without directly informing the implementation. Direct implementations of ideas from logic—theorem-proving and model-construction techniques—are used in AI, but the AI theorists who rely on logic to model their problem areas are free to use other implementation techniques as well. Thus, in [Moore, 1995b, Chapter 1], Robert C. Moore distinguishes three uses of logic in AI; as a tool of analysis, as a basis for knowledge representation, and as a programming language.

A large part of the effort of developing limited-objective reasoning systems goes into the management of large, complex bodies of declarative information. It is generally recognized in AI that it is important to treat the representation of this information, and the reasoning that goes along with it, as a separate task, with its own research problems.

The evolution of expert systems illustrates the point. The earliest expert systems, such as Mycin (a program that reasons about bacterial infections, see [Buchanan and Shortliffe, 1984]), were based entirely on large systems of procedural rules, with no separate representation of the background knowledge—for instance, the taxonomy of the infectious organisms about which the system reasoned was not represented.

Later generation expert systems show a greater modularity in their design. A separate knowledge representation component is useful for software engineering purposes—it is much better to have a single representation of a general fact that can have many different uses, since this makes the system easier to develop and to modify. And this design turns out to be essential in enabling these systems to deliver explanations as well as mere conclusions.[4]

1.3. Knowledge Representation

In response to the need to design this declarative component, a subfield of AI known as *knowledge representation* emerged during the 1980s. Knowledge representation deals primarily with the representational and reasoning challenges of this separate component. The best place to get a feel for this subject is the proceedings of the meetings that are now held every other year: see [Brachman *et al.*, 1989, Allen *et al.*, 1991, Nebel *et al.*, 1992, Doyle *et al.*, 1994, Aiello *et al.*, 1996, Cohn *et al.*, 1998, Cohn *et al.*, 2000, Fensel *et al.*, 2002].

Typical articles in the proceedings of the KR and Reasoning conferences deal with the following topics.

[4]See [Stefik, 1995] for general background on expert systems. For information concerning explanation, see [Clancey, 1983, Moore, 1995a].

1. Topics in logical theory and the theory of computation, including
    (a) Nonmonotonic logic
    (b) Complexity theory
2. Studies in application areas, including
    (a) Temporal reasoning
    (b) Formalisms for reasoning about planning, action and change
    (c) Metareasoning
    (d) Reasoning about context
    (e) Reasoning about values and desires
    (f) Reasoning about the mental states of other agents, and especially about knowledge and belief
    (g) Spatial reasoning
    (h) Reasoning about vagueness
3. Studies in application techniques, including
    (a) Logic programming
    (b) Description logics
    (c) Theorem proving
    (d) Model construction
4. Studies of large-scale applications, including
    (a) Cognitive robotics
    (b) Merging, updating, and correcting knowledge bases

These topics hardly overlap at all with the contents of the *Journal of Symbolic Logic,* the principal research archive for mathematical logic. But there is substantial overlap in theoretical emphasis with *The Journal of Philosophical Logic*, where topics such as tense logic, epistemic logic, logical approaches to practical reasoning, belief change, and vagueness account for a large percentage of the contributions. Very few *JPL* publications, however, deal with complexity theory or with potential applications to automated reasoning.

1.4. Philosophical Logic

I do not know of a good history of philosophical logic. In fact, the distinction between mathematical and philosophical logic may well be incidental in relation to the overall goals of the subject, since technical rigor and the use of mathematical methods seem to be essential in all areas of logical research. However, the distinction between the two subfields has been magnified by differences in the sorts of professional training that are available to logicians, and by the views of individuals on what is important for the field. The statement of policy presented in Volume 1, no. 1 of the *Journal of Symbolic Logic* (1936) lists bringing together the mathematicians and philosophers working in logic among the goals of the new journal. Probably at this time both the mathematicians and the philosophers shared a sense that their subject was considered to be somewhat marginal by their colleagues, and may have felt a primary loyalty to logic as a subject rather than to any academic discipline. Articles in the first volume of the *JSL* were divided about equally between professional mathematicians and philosophers, and the early volumes of the *JSL* do not show any strong differences between the two groups as to topic.

This situation changed in the 1960s. The 1969 volume of the *JSL* contained 39 articles by mathematicians, and only nine by philosophers. By the early 1970s, many philosophers felt that philosophical papers on logic were unlikely to be accepted by the *JSL,* and that if they were accepted they were unlikely to be read by philosophers. At this point, the goals of the two groups had diverged considerably. Mathematicians were pursuing the development of an increasingly technical and complex body of methods and theorems. Many philosophers felt that this pursuit was increasingly irrelevant to the goal of illuminating philosophical issues. These divisions let to the founding of the *Journal of Philosophical Logic* in 1972. The list of sample topics in the first issue included;

1. Contributions to branches of logical theory directly related to philosophical concerns, such as inductive logic, modal logic, deontic logic, quantum logic, tense logic, free logic, logic of questions, logic of commands, logic of preference, logic of conditionals, many-valued logic, relevance logics;
2. Contributions to philosophical discussions that utilize the machinery of formal logic ...;
3. Discussions of philosophical issues relating to logic and the logical structure of language, ... ;
4. Philosophical work relating to the special sciences, ...

Most of the articles over the subsequent 28 years of the *JPL* belong to the first of these four categories. But the description with which this list begins is not particularly illuminating: why should these particular topics be of interest to philosophers? I believe that the most important feature they share is a sense that despite successes in formalizing areas of mathematical logic, the scope of logic remained severely limited. There are unsolved problems in formalizing the nonmathematical sciences that seem to require thinking through new and different logical issues (quantum logic and the logic of induction, for instance). The remaining topics cover a part, at least, of the even more pressing problems involved in extending logical theory to nonscientific reasoning. The dominant goal, then, of philosophical logic is the extension of logical methods to nonmathematical reasoning domains. This goal has a theoretical dimension if (as many philosophical logicians seem to feel) it requires reworking and extending logical formalisms.

The development and testing of applications (applications such as the problem of formalizing the reasoning involved in getting to the airport, that was posed as a challenge in [McCarthy, 1959]—see Section 2.2, below) doesn't even appear as a category in the list of *JPL* topics, and in fact most of the philosophical logic literature is theoretical.

1.5. Logic in AI and Philosophical Logic

The rough comparison in Section 1.3 of the contents of the main publications for research in logical AI and philosophical logic suggests the following picture. Theoretical work in logical

AI and in philosophical logic overlap to a large extent. Both are interested in developing nonmetamathematical applications of logic, and the core topics are very similar. This overlap is due not only to commonality of interest, but to direct influence of philosophical logic on logical AI; there is ample evidence, as we will see, that the first generation at least of AI logicists read and were influenced by the literature in philosophical logic.

Since that point, the specialties have diverged. New logical theories have emerged in logical AI (nonmonotonic logic is the most important example) which are familiar only to a subcommunity of the philosophical logicians. Other differences are due to the AI community's interest in the theoretical analysis of algorithms and, of course, with their sense of the importance of implementations. Some have to do with the emerging development in computer science of ambitious applications using unprecedentedly large bodies of logical axioms. The sheer size of these applications produces new problems and new methodologies. And other differences originate in the interest of philosophical logicians in some topics (metaphysical topics, for instance) that are primarily inspired by purely philosophical considerations.

Concern for applications can be a great influence on how research is carried out and presented. The tradition in philosophical logic predates applications in automated reasoning, and to this day remains relatively uninterested in such applications. The methodology depends on intuitions, but without any generally accepted methodology for articulating and deploying these intuitions. And the ideas are illustrated and informed by artificial, small-scale examples.[5] In general, the philosophical literature does not deal with implementability or efficiency of the reasoning, or indeed with any features of the reasoning process. And it is hard to find cases in which the philosophical theories are illustrated or tested with realistic, large-scale reasoning problems.

These differences, however, are much more a matter of style than of substance or of strategic research goals. It is difficult to think through the details of the reasoning process without the computational tools to make the process concrete, and difficult to develop large-scale formalizations of reasoning problems without computational tools for entering, testing, and maintaining the formalizations. Because the core theoretical topics (modal, conditional and temporal logic, belief revision, and the logic of context) are so similar, and because the ultimate goal (the formalization of nonmathematical reasoning) is the same, I think of logic in AI as a continuous extension of the philosophical logic tradition.

The early influence of philosophical logic on logic in AI was profound. The bibliography of [McCarthy and Hayes, 1969], one of the most influential early papers in logical AI, illustrates the point well. There are 58 citations in the bibliography. Of these, 35 refer to the philosophical logic literature. (There are 17 computer science citations, one mathematical logic citation, one economics citation, and one psychology citation.) This paper was written at a time when there were hardly any references to logical AI in the computer science literature. Naturally, as logical AI has matured and developed as a branch of computer science, the proportion of cross-disciplinary citations has decreased. A sampling of articles from the first Knowledge Representation conference, [Brachman *et al.*, 1989], held in 1989, shows only 12 philosophical logic citations out of a total of 522 sampled citations; a sampling of articles from [Cohn *et al.*, 1998], held in 1998, shows 23 philosophical logic citations out of a total

[5] For a good example of the use of these intuitions to motivate a system of logic, see the extended argument in [Hintikka, 1962] that the modal logic **S4** is the correct logic of belief.

of 468 sampled.[6]

Despite the dramatic decrease in quantity of explicit citations, the contemporary literature in logical AI reflects an indirect acquaintance with the earlier literature in philosophical logic, since many of the computational papers that are explicitly cited in the modern works were influenced by this literature. Of course, the influence becomes increasingly distant as time passes, and this trend is accelerated by the fact that new theoretical topics have been invented in logical AI that were at best only dimly prefigured in the philosophical literature.

Although philosophical logic is now a relatively small field in comparison to logical AI it remains a viable area of research, with new work appearing regularly. But references to contemporary research in philosophical logic are rare in the AI literature. Similarly, the papers currently published in *The Journal of Philosophical Logic*, at least, do not show much influence from AI.[7] In Europe, the lines are harder to draw between professional divisions among logicians: some European journals, especially the *Journal of Logic, Language, and Information*, are successful in maintaining a focus in logic while attracting authors from all the disciplines in which logic is represented.

1.6. The Role of Artificial Intelligence in Logic

The importance of applications in logical AI, and the scale of these applications, represents a new methodology for logic—one that would have been impossible without mechanized reasoning. This methodology forces theoreticians to think through problems on a new scale and at a new level of detail, and this in turn has a profound effect on the resulting theories. The effects of this methodology will be illustrated in the sections below, dealing with various topics in logical AI. But the point is illustrated well by reasoning about action and change. This topic was investigated in the philosophical literature. Reasoning about change, at least, is part of tense logic, and the consequences of action are investigated in the literature on "seeing to it that"; see, for instance, [Belnap, Jr., 1996]. The latter theory has no very robust account of action. The central construct is a variation on a branching-time modality of the sort that has been familiar since [Prior, 1967]. Although it represents an interesting development in philosophical logic, the scale of the accomplishment is very different from the research tradition in logical AI reported in Section 4, below. The formalisms in this tradition not only support the formalization of complex, realistic planning problems, but provide entirely new insights into reasoning about the causal effects of actions, the persistence of states, and the interactions between actions and continuous physical processes. Developments such as this would have been impossible without the interactions between the logical theories and large-scale, practical applications in automated planning.

In [Carnap, 1955], Rudolf Carnap attempted to clarify intensional analyses of linguistic meaning, and to justify from a methodological point of view, by imagining how the analysis could be applied to the linguistic usage of a hypothetical robot. Carnap hoped that the fact that we could imagine ourselves to know the internal structure of the robot would help to

[6] The submissions to the 1989 conference were unclassified as to topic; I sampled every other article, a total of 522. The 1989 conference divided its contributed articles into 26 topical sessions; I sampled the first paper in each of these sessions.

[7] In the decade from 1990 to 1999 I counted one *JPL* publication by an AI researcher, [Boutilier, 1996], and five papers showing some AI influence; all of these dealt with nonmonotonic logic.

make the case for an empirical science of semantics more plausible. This hope proved to be unjustified; the philosophical issue that concerned Carnap remains controversial to this day, and thought experiments with robots have not proved to be particularly rewarding in addressing it. Real robots, though, with real applications,[8] are a very different matter. Though it is hard to tell whether they will prove to be helpful in clarifying fundamental philosophical problems, they provide a laboratory for logic that is revolutionary in its potential impact on the subject. They motivate the development of entirely new logical theories that I believe will prove to be as important for philosophy as the fundamental developments of the late nineteenth century proved to be.

The emergence of separate mathematical and philosophical subspecialties within logic was not an entirely healthy thing for the field. The process of making mathematical logic rigorous and of demonstrating the usefulness of the techniques in pursuing mathematical ends that was pursued so successfully in the first half of the twentieth century represents a coherent refinement of logical methodology. All logicians should be pleased and proud that logic is now an area with a body of results and problems that is as substantial and challenging as those associated with most areas of mathematics.

But these methodological advances were gained at the expense of coverage. In the final analysis, logic deals with reasoning—and relatively little of the reasoning we do is mathematical, while almost all of the mathematical reasoning that nonmathematicians do is mere calculation. To have both rigor and scope, logic needs to keep its mathematical and its philosophical side united in a single discipline. In recent years, neither the mathematical nor the philosophical professions—and this is especially true in the United States—have done a great deal to promote this unity. But the needs of Computer Science, provide strong unifying motives. The professional standards for logical research in Computer Science certainly require rigor, but the field also puts its practitioners into contact with reasoning domains that are not strictly mathematical, and creates needs for innovative logical theorizing.

The most innovative and ambitious area of Computer Science, in terms of its coverage of reasoning, and the one that is closest in spirit to philosophical logic, is AI. This chapter will attempt to provide an introduction, for outsiders who are familiar with logic, to the aspects of AI that are closest to the philosophical logic tradition. This area of logic deserves, and urgently needs, to be studied by historians. But I am not a historian, and this document does not pretend to be a history.

2. John McCarthy and Common Sense Logicism[9]

2.1. Logical AI

The most influential figure in logical AI is John McCarthy. McCarthy is one of the founders of AI, and has consistently advocated a research methodology that uses logical techniques to formalize the reasoning problems that AI needs to solve. All but the most recent work

[8]This includes robots (or "softbots") that navigate artificial environments such as the Internet or virtual worlds as well as embodied robots that navigate the physical world.

[9]I was surprised at first to hear the AI community refer to its logical advocates as logicists. On reflection, it seems to me much better to think of logicist projects in this general sense, as proposals to apply what Alonzo Church called "the logistic method" in seeking to understand reasoning in various domains. It is far too restrictive to narrowly associate logicism with Frege's program.

in McCarthy's research program can be found in [Lifschitz, 1990a], which also contains an introduction to McCarthy's work, [Lifschitz, 1990b]; for additional historical background, see [Israel, 1991].

McCarthy's methodological position has not changed substantially since it was first articulated in [McCarthy, 1959] and elaborated and amended in [McCarthy and Hayes, 1969]. The motivation for using logic is that—even if the eventual implementations do not directly and simply use logical reasoning techniques like theorem proving—a logical formalization helps us to understand the reasoning problem itself. The claim is that without an understanding of what the reasoning problems are, it will not be possible to implement their solutions. Plausible as this Platonic argument may seem, it is in fact controversial in the context of AI; an alternative methodology would seek to learn or evolve the desired behaviors. The representations and reasoning that this methodology would produce might well be too complex to characterize or to understand at a conceptual level.

From [McCarthy and Hayes, 1969], it is clear that McCarthy thinks of his methodology for AI as overlapping to a large extent with traditional philosophy, but adding to it the need to inform the design of programs capable of manifesting general intelligence. This idea is not uncongenial to some philosophers (see, for instance, [Carnap, 1956, pp. 244–247] and [Pollock, 1995]), and I personally believe that ideas that logical AI is potentially of great value for philosophy. In practice, however, the actual theories that have emerged from McCarthy's methodology are influenced most strongly by work in philosophical logic, and the research tradition in logical AI represents a more or less direct development of this work, with some changes in emphasis. This review will concentrate on logical AI in relation to philosophical logic, without further comment on relations to philosophy in general or to the feasibility of developing human-level intelligent systems.

2.2. The Formalization of Common Sense

McCarthy's long-term objective is to formalize *common sense reasoning*, the prescientific reasoning that is used in dealing with everyday problems. An early example of such a problem, mentioned in [McCarthy, 1959], is getting from home to the airport. Other examples include:

1. *Narrative understanding.* The reasoning involved in reconstructing implicit information from narratives, such as sequencing of eventualities, and inferred causal connections.
2. *Diagnosis.* For instance, detecting faults in physical devices.
3. *Spatial Reasoning.* For instance, reasoning about the parts of rigid bodies and their shapes, and their relation to the shape of the whole.
4. *Reasoning about the attitudes of other agents.* For instance, making informed guesses about the beliefs and desires of other agents, not from "keyhole observation" but from conversational clues of the sort that could be obtained in a brief, interactive interview.

Stated baldly, the goal of formalizing common sense would probably seem outrageous to most philosophers, who are trained to think of common sense as rather elusive. But whether or not the ultimate goal is appropriate and achievable, the specific formalization projects that have emerged from this program have been successful in several ways. They have succeeded in breaking new territory for logic by extending the scope of the reasoning problems to which logical techniques can be successfully applied. They have demonstrated that logical techniques can be an important part of the solutions to specific AI problems—planning is the most successful of these, but some success has been achieved in other areas as well.[10] They form the basis of one approach to developing complete, autonomous agents.[11] And they have illuminated many specific forms of nonscientific reasoning—for instance, qualitative reasoning about the behavior of physical devices.[12]

3. Nonmonotonic Reasoning and Nonmonotonic Logics

3.1. Nonmonotonicity

Aristotle believed that most reasoning, including reasoning about what to do and about sublunary natural phenomena, dealt with things that hold "always or for the most part." But Aristotelian logic deals only with patterns of inference that hold without exception. We find at the very beginning of logic a discrepancy between the scope of logical theory and common sense reasoning. Nonmonotonic logic is the first sustained attempt within logical theory to remedy this discrepancy. As such, it represents a potential for a sweeping expansion of the scope of logic, as well as a significant body of technical results.

The consequence relations of classical logics are monotonic. That is, if a set Γ of formulas implies a consequence C then a larger set $\Gamma \cup \{A\}$ will also imply C. A logic is *nonmonotonic* if its consequence relation lacks this property. *Preferred models* provide a general way to induce a nonmonotonic consequence relation. Invoke a function that for each Γ produces a subset $\mathcal{M}_\Gamma$ of the models of Γ; in general, we will expect $\mathcal{M}_\Gamma$ to be a proper subset of the models of Γ. We then say that Γ implies C if C is satisfied by every model in $\mathcal{M}_\Gamma$. As long as we do not suppose that $\mathcal{M}_{\Gamma \cup \{A\}} \subseteq \mathcal{M}_\Gamma$, we can easily have an implication relation between Γ and C without imposing this relation on supersets of Γ.[13]

This model theoretic behavior corresponds to expectation-guided reasoning, where the expectations allow certain cases to be neglected. Here is an important difference between common sense and mathematics. Mathematicians are trained to reject a proof by cases unless the cases exhaust all the possibilities; but typical instances of common sense reasoning neglect some alternatives. In fact, it is reasonable to routinely ignore outlandish possibilities.

[10] Data integration is one such area. See [Levy, 2000]. Large-scale knowledge representation is another. See [Lenat and Guha, 1989].

[11] See [Reiter, 2001] for an extended contribution to *cognitive robotics*, with references to some of the other literature in this area. Reiter's book also contains self-contained chapters on the Situation Calculus and the problems of formalizing reasoning about action and change. I recommend these chapters to anyone wishing to follow up on the topics discussed in Section 4. Another extended treatment of action formalisms and issues is [Shanahan, 1997]

[12] But much of the work in this last area has not made heavy use of logical techniques. Qualitative physics and the formalization of other forms of qualitative reasoning is an independent specialty in AI, different in many ways from logical AI. But the two specialties have certainly influenced each other. For information concerning qualitative reasoning, consult [Forbus, 1988, Weld and de Kleer, 1990, Kuipers, 1993].

[13] For further details concerning this approach, see Section 3.4, below.

Standing in my kitchen in California, wondering if I have time to wash my dishes before leaving for work, I do not take the possibility of an earthquake into account.

There seem to be many legitimate reasons for neglecting certain cases in common sense reasoning. A qualitative judgment that the probability of a case is negligible is one reason. But, for instance, in a planning context it may be reasonable to ignore even nonnegligible probabilities, as long as there is no practical point in planning on these cases.

The motivations for nonmonotonicity seem to involve a number of complex factors; probability (perhaps in some qualitative sense), normality, expectations that are reasonable in the sense that one can't be reasonably blamed for having them, mutual acceptance, and factors having to do with limited rationality. As far as I know, no one has succeeded in disentangling and clarifying these motivating considerations. In the early stages of its emergence in logical AI, many researchers seem to have thought of nonmonotonic reasoning as a general method for reasoning about uncertainty; but by the end of the 1980s, implementations of fully quantitative probabilistic reasoning were not only possible in principle, but were clearly preferable in many sorts of applications to methods involving nonmonotonic logic. A plausible and realistic rationale for nonmonotonic logic has to fit it into a broader picture of reasoning about uncertainty that also includes probabilistic reasoning.[14]

3.2. Historical Motivations

Three influential papers on nonmonotonic logic appeared in 1980: [McCarthy, 1980, McDermott and Doyle, 1980, Reiter, 1980]. In each case, the formalisms presented in these papers were the result of a gestation period of several years or more. To set out the historical influences accurately, it would be necessary to interview the authors, and this I have not done. However, there seem to have been two motivating factors: strategic considerations having to do with the long-range goals of AI, and much more specific, tactical considerations arising from the analysis of the reasoning systems that were being deployed in the 1970s.

Section 2.2 drew attention to McCarthy's proposed goal of formalizing common sense reasoning. The brief discussion above in Section 3.1 suggests that monotonicity may be an obstacle in pursuing this goal. An additional motive was found in [Minsky, 1974], which was widely read at the time. This paper presents an assortment of challenges for AI, focusing at the outset on the problem of natural language understanding.[15] Minsky advocates frame-based knowledge representation techniques[16] and (conceiving of the use of these representations as an alternative to logic), he throws out a number of loosely connected challenges for the logical approach, including the problem of building large-scale representations, of rea-

[14] John McCarthy makes a similar point, illustrating it with an example, in [McCarthy, 1993a].

[15] This very difficult and not particularly well-defined problem was very much on the minds of many AI researchers in the area that later became knowledge representation, but it has not proved to be a productive focus for the field. Natural language interpretation has developed into a separate field, that is largely concerned with less sweeping problems, such as automated speech-to-speech discourse, data mining, and text summarization. Logical techniques have been used with some success in this area, but it is fair to say that natural language interpretation has not been the best showcase for logical ideas. Even the problem of providing an adequate semantic interpretation of generic constructions—a natural application of nonmonotonic logic—has turned out to be problematic. See [Krifka *et al.*, 1995] for a general discussion of the issues.

[16] This use of the word 'frame' is unconnected to the use of the term in the "frame problem," and is not to be confused with that problem.

soning efficiently, of representing control knowledge, and of providing for the flexible revision of defeasible beliefs. In retrospect, I think most AI researchers would agree that these problems are general challenges to any research program in AI (including the one Minsky himself advocated at the time) and that logical techniques are an important element in addressing some, perhaps all, of the issues. (For instance, a well structured logical design can be a great help in scaling up knowledge representation.)

Minsky apparently intended to provide general argument against logical methods in AI, but [McCarthy, 1980, McDermott and Doyle, 1980] interpret it as a challenge that can be met by developing logics that lack the monotonicity property. Perhaps unintentionally, Minsky's paper seems to have provided some incentive to the nonmonotonic logicians by stressing monotonicity as a source of the alleged shortcomings of logic. In fact, the term 'monotonicity' apparently makes its first appearance in print in this paper.

The development of nonmonotonic logic also owes a great deal to the applied side of AI. In fact, the need for a nonmonotonic analysis of a number of AI applications was as persuasive as the strategic considerations urged by McCarthy, and in many ways more influential on the shape of the formalisms that emerged. Here, I will mention three such applications that appear to have been important for some of the early nonmonotonic logicians: belief revision, closed-world reasoning, and planning.

Belief Revision. [Doyle, 1979] provides an analysis and algorithm for a "truth maintenance system". The TMS answers a general need, providing a mechanism for updating the "beliefs" of knowledge bases. The idea of the TMS is to keep track of the support of beliefs, and to use the record of these support dependencies when it is necessary to revise beliefs. In a TMS, part of the support for a belief can consist in the *absence* of some other belief. This introduces nonmonotonicity. For instance, it provides for defaults; that Wednesday is the default day for scheduling a meeting means the belief that the meeting will be on Wednesday depends on the *absence* of the belief that it will not be on Wednesday.

The TMS algorithm and its refinements had a significant impact on AI applications, and this created the need for a logical analysis. (In even fairly simple cases, it can be hard in the absence of analytic tools to see what consequences a TMS should deliver.) This provided a natural and highly specific challenge for those seeking to develop a nonmonotonic logic. The TMS also provided specific intuitions: the idea that the key to nonmonotonicity has to do with inferences based on unprovability was important for the modal approaches to nonmonotonic logic and for default logic. And the TMS's emphasis on interactions between arguments began a theme in nonmonotonic logic that remains important to this day. (See the discussion of argument-based approaches, in Section 3.4, below.)

Closed-world reasoning. The study of databases belongs to computer science, but not to AI. But one of the research paradigms in the scientific analysis of databases uses logical models of the representations and reasoning (see [Minker, 1997] for a recent survey of the field), and this area has interacted often with logical AI. The deductive database paradigm was taking shape at about the same time that many AI researchers were thinking through the problems of nonmonotonic logic, and provided several specific examples of nonmonotonic reasoning that called for analyses. Of these, perhaps the most important is the *closed-world assumption,* according to which—at least as far as simple facts are concerned, represented in the database as positive or negative literals—the system assumes that it knows all that there is to be known. It is the closed world assumption that justifies a negative answer to a query

‘Is there a direct flight from Detroit to Bologna?’ when the system finds no such flight in its data. This is another case of inference from the absence of a proof; a negative is proved, in effect, by the failure of a systematic attempt to prove the positive. This idea, which was investigated in papers such as [Clark, 1978, Reiter, 1978], also provided a challenge for nonmonotonic logics, as well as specific intuitions—note that again, the idea of inference rules depending on the absence of a proof is present here.

Planning. The need for inertial defaults in temporal reasoning—defaults to the effect that things will stay put in the absence of a reason for them to change—arises in attempting to formalize the reasoning needed in planning. This application (the apparent need for a nonmonotonic logic in developing an economical formal solution to the frame problem) provided another specific formal need. One of the earliest attempts to formalize nonmonotonic reasoning, [Sandewall, 1972], addresses this problem. Inertial defaults are an especially important and instructive case study; I will say no more about them here, since they are discussed in detail in Section 4.4, below.

3.3. The Earliest Formalisms

The three 1980 papers mentioned at the beginning of Section 3.2 represent three approaches to nonmonotonic logic that remain important subfields to this day: *circumscription* (McCarthy), *modal approaches* (Doyle and McDermott) and *default logic* (Reiter).

In [McCarthy, 1993a], McCarthy urges us, when considering the early history of circumscription, to take into account a group of three papers: [McCarthy, 1987, McCarthy, 1980, McCarthy, 1986]. The first paper connects the strategic ideas of [McCarthy and Hayes, 1969] with the need for a nonmonotonic logic, and sketches the logical ideas of *domain circumscription*, which is now classified as the simplest case of circumscription. The second paper provides more thorough logical foundations, and introduces the more general and powerful *predicate circumscription* approach. The third paper concentrates on developing techniques for formalizing challenging common sense examples.

All forms of circumscription involve restricting attention to models in which certain sets are minimized; for this reason, circumscription can be grouped with the preferred models approaches to nonmonotonicity: see Section 3.4, below. McCarthy's formalism is fairly conservative; though it raises interesting logical issues in higher-order logic and complexity, it uses familiar logical frameworks. And much of the focus is on the development of formalization techniques. The other varieties of nonmonotonic logic, including default logic and the modal nonmonotonic logics, raise issues of the sort that are familiar to philosophical logicians, having to do with the design of new logics, the systematic investigation of questions concerning validity, and managing the proliferation of logics.

As the discussion above of truth maintenance indicated, it is very natural to think of nonmonotonic inferences as being *hedged*. That is, a nonmonotonic inference may require not merely the presence of a set of proved conclusions, but the *absence* of certain other conclusions. The general form of such a rule is:

> **DR** In the presence of $\{A_1, \ldots, A_n\}$
> and in the absence of $\{B_1, \ldots, B_n\}$,
> conclude C.

An important special case of **DR** is a *normal default*, a simple rule to the effect that C holds by default, conditionally on A. This can be formalized by taking the condition that must be absent to simply be the negation of the conclusion.

NDR In the presence of $\{A_1, \ldots, A_n\}$
and in the absence of $\neg C$,
conclude C.

At first sight, it is somewhat perplexing how to formalize this notion of nonmonotonic inference, since it seems to require a circular definition of provability that can't be replaced with an inductive definition, as in the nonmonotonic case. The difficulty with the early theory of [Sandewall, 1972] is that it does not address this difficulty successfully. [McDermott and Doyle, 1980, Reiter, 1980] use fixpoint definitions to solve the problem. In both cases, the logical task is (1) to develop a formalism in which rules like **DR** can be expressed, and (2) to define the relation between a theory DT (which may incorporate such rules) and the theories E which could count as reasonable consequences of DT. In the terminology that later became standard, we need to define the relation between a theory DT and its *extensions*.

In retrospect, we can identify two sorts of approaches to nonmonotonic logic: those based on *preference* and those based on *conflict*. Theories of the first sort (like circumsctription) involve a relatively straightforward modification of the ordinary model-theoretic definition of logical consequence that takes into account a preference relation over models. Theories of the second sort (like default logic) involve a more radical rethinking of logical consequence. The possibility of multiple extensions—different possible coherent, inferentially complete conclusion sets that can be drawn from a single set of premises—means that we have to think of logical consequence not as a function taking a set of axioms into its logical closure, but as a *relation* between a set of axioms and alternative logical closures. Since logical consequence is so fundamental, this represents a major theoretical departure. With multiple extensions, we can still retrieve a consequence relation between a theory and a formula in various ways, the simplest being to say that DT nonmonotonically implies C if C is a member of every extension of DT. Still, the conflict-based account of consequence provides a much richer underlying structure than the preferential one.

Reiter approaches the formalization problem conservatively. Nonmonotonicity is not expressed in the language of default logic, which is the same as the language of first-order logic. But a theory may involve a set of *default rules*—rules of the form **DR**. [Reiter, 1980] provides a fixpoint definition of the extensions of such a theory, and develops the theoretical groundwork for the approach, proving a number of the basic theorems.

Of these theorems, I mention one in particular, which will be used in Section 4.5, in connection with the Yale Shooting Anomaly. The idea is to take a conjectured extension (which will be a set T^*) and to use this set for consistency checks in a proof-like process that applies default rules in $\langle W, D\rangle$ successively to stages that begin with W.

We define a default proof process $T_0, T_1, \ldots$ for W, D, relative to T^*, as follows.

Let $T_0 = W$.

If no default rule in D is nonvacuously applicable to T_i relative to T^*, then $T_{i+1} = \mathrm{Th}_{\mathrm{FOL}}(T_i)$.

Otherwise, choose some default rule

$$\frac{A \,:\, B_1, \ldots, B_n}{C}$$

that is nonvacuously applicable to T_i relative to T^*, and let

$$T_{i+1} = \mathrm{Th}_{\mathrm{FOL}}(T_i \cup \{C\})$$

In other words, as long as we can nonvacuously close the stage we are working on under an applicable default, we do so; otherwise, we do nothing.

A theorem of Reiter's says that, under these circumstances:

> T is an extension of $\langle W, D\rangle$ if and only if there is a proof process $T_0, T_1, \ldots$ for W, D, relative to T, such that
>
> $$T = \bigcup_{i=0}^{\infty} T_i.$$

Thus, we can show that T is an extension by (1) using T for consistency checks in a default reasoning process from $\langle W, D\rangle$, (2) taking the limit T' of this process, and (3) verifying that in fact $T' = T$.

The modal approach represents a "higher level of nonmonotonic involvement" than default logic. The unprovability construct is represented explicitly in the language, by means of a modal operator L informally interpreted as 'provable' (or, as in [McDermott and Doyle, 1980], by the dual of this operator).[17] Although McDermott and Doyle's terminology is different from Reiter's, the logical ideas are very similar—the essence of their approach, like Reiter's, is a fixpoint definition of the extensions of a nonmonotonic logic. Incorporating nonmonoticity in the object language creates some additional complexities, which in the early modal approach show up mainly in proliferation of the logics and difficulties in evaluating the merits of the alternatives. As better foundations for the modal approach emerged, it became possible to prove the expected theorems concerning equivalence of modal formalisms with default logic.[18]

Reiter's paper [Reiter, 1980] appears to have developed primarily out of tactical considerations. The earlier paper [Reiter, 1978] is largely concerned with providing an account of database queries. Unlike the other seminal papers in nonmonotonic logic, Reiter's shows specific influence from the earlier and independent work on nonmonotonicity in logic programming—the work seems to have been largely inspired by the need to provide logical foundations for the nonmonotonic reasoning found in deductive databases. Doyle and McDermott's paper shows both strategic and tactical motivation—citing the earlier literature

[17] The analogy to modal logics of provability inspired by Gödel's work ([Boolos, 1993]) has, of course, been recognized in later work in nonmonotonic logic. But it has not been a theme of major importance.

[18] See [Konolige, 1988].

in logicist AI, it motivates nonmonotonic logic as part of a program of modeling common sense rationality. But the theory is also clearly influenced by the need to provide a formal account of truth maintenance.

3.4. Approaches to Nonmonotonic Logic

Nonmonotonic logic is a complex, robust research field. Providing a survey of the subject is made difficult by the fact that there are many different foundational paradigms for formalizing nonmonotonic reasoning, and the relations between these paradigms is not simple. An adequate account of even a significant part of the field requires a something like a book-length treatment. A number of books are available, including [Antoniou, 1997, Besnard, 1992, Brewka, 1991, Brewka *et al.*, 1997, Lukaszewicz, 1990, Marek and Truszczyński, 1994, Schlechta, 1997]. Two collections are especially useful: [Ginsberg, 1987] and [Gabbay *et al.*, 1994]. The former is a useful source for readers interested in the early history of the subject, and has an excellent introduction. The handbook chapters in [Gabbay *et al.*, 1994] provide overviews of important topics and approaches. My current recommendation for readers interested in a quick, readable introduction to the topic would be [Brewka *et al.*, 1997] and self-selected chapters of [Gabbay *et al.*, 1994]. I will rely on these references for technical background, and will concentrate on intellectual motivation, basic ideas, and potential long-term significance for logic.

Preference Semantics At the outset in Section 3.1, I mentioned how preferred models could be used to characterize a nonmonotonic consequence relation. This general model theory of nonmonotonicity emerged in [Shoham, 1988] five years after the work discussed in Section 3.2, and represents a much more general and abstract approach.

Preferential semantics rely on a function S taking a set K of models into a subset $S(K)$ of K. The crucial definition of *preferential entailment* stipulates that A is a (nonmonotonic) consequence of Γ if every model M of $S(\llbracket\Gamma)\rrbracket)$ implies A. Shoham's theory is based on a partial order $\preceq$ over models: $S(K)$ can then be characterized as the set of models in K that are $\preceq$-minimal in K. To ensure that no set can preferentially entail a contradiction unless it classically entails a contradiction, infinite descending $\preceq$ chains need to be disallowed.

This treatment of nonmonotonicity is similar to the earlier modal semantic theories of conditionals—the similarities are particularly evident using the more general theories of conditional semantics, such as the one presented in [Chellas, 1975]. Of course, the consequence relation of the classical conditional logics is monotonic, and conditional semantics uses possible worlds, not models. But the *left-nonmonotonicity* of conditionals (the fact that $A \;\Box\!\!\rightarrow\; C$ does not imply $[A \wedge B] \;\Box\!\!\rightarrow\; C$) creates issues that parallel those in nonmonotonic logics. Early work in nonmonotonic logic does not seem to be aware of the analogy with conditional logic. But the interrelations between the two have become an important theme more recently; see, for instance, [Alcourrón, 1995, Arlo-Costa and Shapiro, 1992, Asher, 1995, Benferat *et al.*, 1997, Boutilier, 1992, Delgrande, 1998, Gabbay, 1995, Gärdenfors and Makinson, 1994, Pearl, 1994].

Preference semantics raises an opportunity for formulating and proving representation theorems relating conditions over preference relations to properties of the abstract consequence relation. This line of investigation began with [Lehmann and Magidor, 1992].

Modal and epistemic theories Neither Doyle or McDermott pursued the modal approach much beyond the initial stages of [McDermott and Doyle, 1980, McDermott, 1982]. With a helpful suggestion from Robert Stalnaker (see [Stalnaker, 1993]), however, Robert C. Moore produced a modal theory that improves in many ways on the earlier ideas. Moore gives the modal operator of his system an epistemic interpretation, stressing the interpretation a default rule as one that licenses a conclusion for a reasoning agent unless something that the agent knows blocks the conclusion. In Moore's *autoepistemic logic,* an extension E of a theory T is a superset of T that is *stable,* i.e. that is deductively closed, and that satisfies the following two rules:

(1) If $A \in E$ then $\Box A \in E$;
(2) If $A \notin E$ then $\neg \Box A \in E$;

It is also usual to impose a *groundedness* condition on autoepistemic extensions of T, ensuring that every member of an extension has some reason tracing back to T. Various such conditions have been considered; the simplest one restricts extensions to those satisfying

(3) E is the set of nonmodal consequences of $T \cup \{A : \Box A \in E\} \cup \{\neg \Box A : A \notin E\}$.

Autoepistemic logic remains a popular approach to nonmonotonic logic, in part because of its usefulness in providing theoretical foundations for logic programming. For more recent references, see [Antoniou, 1997, Konolige, 1994, Marek and Truszczyński, 1989, Marek and Truszczyński, 1991, Moore, 1993, Moore, 1995b].

Epistemic logic has inspired other approaches to nonmonotonic logic. Like other modal theories of nonmonotonicity, these use modality to reflect consistency in the object language, and so allow default rules along the lines of **DR** to be expressed. But instead of consistency, these use *ignorance.* See [Halpern and Moses, 1985] and [Levesque, 1987] for variations on this idea. These theories are explained, and compared to other nonmonotonic logics, in [Meyer and van der Hoek, 1995]. In more recent work, Levesque's ideas are systematically presented and applied to the theory of knowledge bases in [Levesque and Lakemeyer, 2000].

3.5. Further Topics

This brief historical introduction to nonmonotonic logic leaves untouched even a number of general topics that might well be of interest to a nonspecialist. These include graph-based and proof-theoretic approaches to nonmonotonic logic, results that interrelate the various formalisms, complexity results, tractable special cases of nonmonotonic reasoning, relations between nonmonotonic and abductive reasoning, relations to probability logics, the logical intuitions and apparent patterns of validity underlying nonmonotonic logics, and the techniques used to formalize domains using nonmonotonic logics. For these and other topics I have to refer the reader to the literature. As a start, I highly recommend the chapters in [Gabbay *et al.*, 1994].

4. Reasoning about Action and Change[19]

4.1. Priorian Tense Logic

Time and temporal reasoning have been associated with logic since the origins of scientific logic with Aristotle. The idea of a logic of tense in the modern sense has been familiar since at least the work of Jan Łukasiewicz (see, for instance, [Łukasiewicz, 1970]), but the shape of what is commonly known as tense logic was standardized by Arthur Prior's work in the 1950s and 1960s: see [Prior, 1956, Prior, 1967, Prior, 1968].[20] As the topic was developed in philosophical logic, tense logic proved to be a species of modal logic; Prior's work was heavily influenced by both Hintikka and Kripke, and by the idea that the truth of tense-logical formulas is relative to *world-states* or temporal stages of the world; these are the tense-theoretic analogues of the timeless possible worlds of ordinary modal logic. Thus, the central logical problems and techniques of tense logic were borrowed from modal logic. For instance, it became a research theme to work out the relations between axiomatic systems and the corresponding model theoretic constraints on temporal orderings. See, for instance, [Burgess, 1984, van Benthem, 1983].

Priorian tense logic shares with modal logic a technical concentration on issues that arise from using the first-order theory of relations to explain the logical phenomena, an expectation that the important temporal operators will be quantifiers over world-states, and a rather distant and foundational approach to actual specimens of temporal reasoning. Of course, these temporal logics do yield validities, such as

$$A \rightarrow PFA,$$

(if A, then it was the case that A was going to be the case), which certainly are intuitively valid. But at most, these can only play a broadly foundational role in accounting for realistic reasoning about time. It is hard to think of realistic examples in which they play a leading part.

This characteristic, of course, is one that modal logic shares with most traditional and modern logical theories; the connection with everyday reasoning is rather weak. Although modern logical techniques do account with some success for the reasoning involved in verifying mathematical proofs and logic puzzles, they do not explain other cases of technical or commonsense reasoning with much detail or plausibility. Even in cases like legal reasoning, where logicians and logically-minded legal theorists have put much effort into formalizing the reasoning, the utility of the results is controversial.

4.2. Planning Problems and the Situation Calculus

Planning problems provide one of the most fruitful showcases for combining logical analysis with AI applications. On the one hand there are many practically important applications of automated planning, and on the other logical formalizations of planning are genuinely helpful in understanding the problems and designing algorithms.

[19] Readers interested in the historical aspects of the material discussed in this section might wish to compare it to [Ohrstrom and Hasle, 1995].

[20] For additional historical background on Prior's work, see [Copeland, 1996].

The classical representation of an AI planning problem, as described in [Amarel, 1968], evidently originates in early work of Herbert Simon's, published in a 1966 CMU technical report, [Simon, 1966]. In such a problem, an agent in an initial world-state is equipped with a set of *actions,* which are thought of as partial functions transforming world-states into world-states. Actions are feasible only in world-states that meet certain constraints (these constraints are now called the "preconditions" of the action). A planning problem then becomes a search for a series of feasible actions that successively transform the initial world-state into a desired world-state.

The *Situation Calculus*, developed by John McCarthy, is the origin of most of the later work in formalizing reasoning about action and change. It was first described in 1969, in [McCarthy, 1983]; the earliest generally accessible publication on the topic is [McCarthy and Hayes, 1969].

Apparently, Priorian tense logic had no influence on [Amarel, 1968]. But there is no important difference between Amarel's world-states and those of Priorian tense logic. The "situations" of the Situation Calculus are these same world-states, under a new name.[21] They resemble possible worlds in modal logic in providing abstract locations that support a consistent and complete collection of truths. As in tense logic, these locations are ordered, and change is represented by the variation in truths from one location to another. The crucial difference between the Situation Calculus and tense logic is that change in the situation is *dynamic*—changes do not merely occur, but occur for a reason.

This difference, of course, is inspired by the intended use of the Situation Calculus: it is meant to formalize Simon's representation of the planning problem, in which a single agent reasons about the scenarios in which a series of actions is performed.[22] In this model, what drives change is the performance of actions, so the fundamental model theoretic relation is the relation

$$\textsc{Result}(\mathrm{a}, \mathrm{s}, \mathrm{s}')$$

between an action a, an initial situation s in which a is performed, and a resulting situation s' immediately subsequent to the performance of the action. Usually (though this is not absolutely necessary) the deterministic assumption is made that s' is unique. In general, actions can be successfully performed only under certain limited circumstances. This could be modeled by allowing for cases in which there is no s' such that $\textsc{Result}(\mathrm{a}, \mathrm{s}, \mathrm{s}')$. But usually, it is assumed that $\textsc{Result}$ is in fact a total function, but that in cases in which s does not meet the "preconditions" of a, there are no restrictions on the s' satisfying $\textsc{Result}(\mathrm{a}, \mathrm{s}, \mathrm{s}')$, so that the causal effects of a will be entirely unconstrained in such cases.

A planning problem starts with a a limited repertoire of actions (where sets of preconditions and effects are associated with each action), an initial situation, and a goal (which can be treated as a formula). A planning problem is a matter of finding a sequence of actions

[21] In retrospect, the term "situation" is not entirely fortunate, since it was later adopted independently and in quite a different sense by the situation semanticists. (See, for instance. [Seligman and Moss, 1996]. In the AI literature, the term "state" is often used interchangeably with "situation", and as far as I can see, without causing any confusion: the connections with physical states, as well as with the more general states of any complex dynamic system are entirely appropriate.

[22] The early versions of the Situation Calculus were meant to be compatible with concurrent cases, i.e., with cases in which there are multiple planning agents, possibly acting simultaneously. But most of the logical analyses have been devoted to the single-agent case.

that will achieve the goal, given the initial situation. That is, given a goal G and initial situation s, the problem will consist of finding a sequence $\mathrm{s}_1, \ldots, \mathrm{s}_n$ of actions which will transform s into a final situation that satisfies G. This means (assuming that RESULT is a function) that G will be satisfied by the situation s_n, where $\mathrm{s}_0 = \mathrm{s}$ and s_{i+1} is the s' such that $\textsc{Result}(\mathrm{a}_{i+1}, \mathrm{s}_i, \mathrm{s}')$. The planning problem is in effect a search for a sequence of actions meeting these conditions. The success conditions for the search can be characterized in a formalism like the Situation Calculus, which allows information about the results of actions to be expressed.

Nothing has been said up till now about the actual language of the Situation Calculus. The crucial thing is how change is to be expressed. With tense logic in mind, it would be natural to invoke a modality like $[\,a\,]A$, with the truth condition

$$\models_{\mathrm{s}} [\,a\,]A \text{ iff } \models_{\mathrm{s}'} A, \text{ where } \textsc{Result}(\mathrm{a}, \mathrm{s}, \mathrm{s}').$$

This formalization, in the style of dynamic logic, is in fact a leading candidate; see Section 4.7, below.

But [McCarthy and Hayes, 1969] deploys a language that is much closer to first-order logic. (This formalization style is characteristic of McCarthy's work; see [McCarthy, 1979].) Actions are treated as individuals. And certain propositions whose truth values can change over time (propositional *fluents*) are also treated as individuals. Where s is a situation and f is a fluent, $\mathit{Holds}(\mathrm{f}, \mathrm{s})$ says that f is true in s.

4.3. Formalizing Microworlds

Since the pioneering work of the nineteenth and early twentieth century logicians, the process of formalizing mathematical domains has largely become a matter of routine. Although (as with set theory) there may be controversies about what axioms and logical infrastructure best serve to formalize an area of mathematics, the methods of formalization and the criteria for evaluating them are relatively unproblematic. This methodological clarity has not been successfully extended to other domains; even the formalization of the empirical sciences presents difficult problems that have not yet been resolved.[23]

The formalization of commonsense reasoning presents an extreme with respect to such methodological difficulties. The work in logical AI has not converged successfully on a solution to this problem. But it has provided the idea of formalizing *microworlds* that represent limited domains of knowledge and reasoning, and work on formalizing these domains has provided some instructive case studies. In addition, there are a few projects that strive for more extensive coverage, as well as some useful methodological ideas. An adequate study of this work would take up a great deal of space. Here, I will only mention some topics and provide some references to the literature.

Temporal reasoning, and in particular reasoning about actions and plans, is the best-developed domain. At least one important methodology will emerge in Section 4.5, below: the development of a library of *scenarios* for testing the adequacy of various formalisms, as the creation of specialized domains like the blocks-world domain (mentioned above, in

[23] Carnap's attempts to formalize dispositional terms and inductive methods are classical examples of the problems that emerge in the formalization of empirical science.

Section 4.2) that serve a laboratories for testing ideas. For more on the blocks world, see [Davis, 1991, Genesereth and Nilsson, 1987]. McCarthy's ideas about *elaboration tolerance* ([McCarthy, 1999]) provide one interesting attempt to provide a criterion for the adequacy of formalizations. Still other important ideas have emerged in the course of formalizing commonsense domains. One is the importance of an explicit *ontology*; see, for instance, [Lenat and Guha, 1989, Fikes, 1996]. Another is the potential usefulness of explicit representations of context; see [Guha, 1991]. Finally, [Davis, 1991] provides many extended examples of formalizations of commonsense domains.

4.4. Prediction and the Frame Problem

To tell whether a plan achieves its goal, you need to see whether the goal holds in the plan's final state. Doing this requires *predictive* reasoning, a type of reasoning that was, as far as I know, entirely neglected in the tense-logical literature. As in mechanics, prediction involves the inference of later states from earlier ones. But (in the case of simple planning problems at least) the dynamics are determined by actions rather than by differential equations. The investigation of this qualitative form of temporal reasoning, and of related sorts of reasoning (e.g., plan recognition, which seeks to infer goals from observed actions, and narrative explanation, which seeks to fill in implicit information in a temporal narrative) is one of the most impressive chapters in the brief history of common sense logicism.

The essence of prediction is the problem of inferring what holds in the situation that ensues from performing an action, given information about the initial situation. I will assume that the agent has complete knowledge about the initial situation—this assumption is usual in classical formalizations of planning.[24]

A large part of the qualitative dynamics that is needed for planning consists in inferring what does *not* change. Take a simple plan to type the word 'cat' using word processing software: my plan is to first enter 'c', then enter 'a', then enter 't'. Part of my confidence in this plan is that the actions are independent: for instance, entering 'a' does not also erase the 'c'. The required inference can be thought of as a form of *inertia*. The *Frame Problem* is the problem of how to formalize the required inertial reasoning.

The Frame Problem was named and introduced in [McCarthy and Hayes, 1969]. Unlike most of the philosophically interesting technical problems to emerge in AI, it has attracted the interest of philosophers; most of the relevant papers, and background information, can be found in [Pylyshyn, 1987, Ford and Pylyshyn, 1996]. Both of these volumes document interactions between AI and philosophy.

The quality of these interactions is discouraging; as a philosopher, I even find it somewhat embarrassing. Like any realistic commonsense reasoning problem, the Frame Problem is open-ended, and can depend on a wide variety of circumstances. If I put $20 in my wallet and go to the store with the wallet in my pocket, I can safely assume that the $20 is still in my wallet. If I leave the $20 on the counter at the store while shopping, I can't safely assume it will be there when I get back. This may account for the temptation that makes some philosophers[25] want to construe the Frame Problem very broadly, so that very soon

[24]For information about planning under uncertainty, see, for instance, [Bacchus *et al.*, 1999, Boutilier *et al.*, 1996, DeJong and Bennett, 1989].

[25]Examples are [Dennett, 1987] and [Fodor, 1987].

it becomes indiscernible from the problem of formalizing general common sense in arbitrary domains. Such a broad construal may serve to introduce speculative discussions concerning the nature of AI, but it loses all contact with the genuine, new logical problems in temporal reasoning that have been discovered by the AI community. It provides a forum for repeating some familiar philosophical themes, but it brings nothing new to philosophy. I find this approach disappointing because I believe that philosophy can use all the help it can get, and that the AI community has succeeded in extending and enriching the application of logic to common sense reasoning in dramatic ways that are highly relevant to philosophy. The clearest account of these developments to be found in the volumes edited by Pylyshyn is [Morgenstern, 1996]. A recent extended treatment can be found in [Shanahan, 1997]; also see [Sandewall, 1994].

The purely logical Frame Problem can be solved using monotonic logic, by simply writing explicit axioms stating what does *not* change when an action is performed. This technique can be successfully applied to quite complex formalization problems.[26] But *nonmonotonic* solutions to the framework have been extensively investigated and deployed; these lead to new and interesting lines of logical development.

Some philosophers ([Fodor, 1987, Lormand, 1996]) have felt that contrived propositions will pose special difficulties in connection with the Frame Problem. As Shanahan points out ([Shanahan, 1997][p. 24]) Fodor's "fridgeon" example is readily formalized in the Situation Calculus and poses no special problems. However, as Lormand suggests, Goodman's examples ([Goodman, 1946]) do create problems if they are admitted as fluents; there will be anomalous extensions in which objects change from green to blue in order to preserve their grueness.

This is one of the few points that I can find in the philosophical literature on the Frame Problem that raises a genuine difficulty for the formal solutions. But the difficulty is peripheral, since the example is not realistic. Recall that fluents are represented as first-order individuals. Although fluents are situation-dependent functions, an axiom of comprehension is certainly not assumed for fluents. In fact, it is generally supposed that the domain of fluents will be a very limited set of the totality of situation-dependent functions; typically, it will be a relatively small finite set of important variables, and will be chosen in particular cases much as a set of variables is chosen in statistical modeling.

I know of no systematic account in the AI literature of how to choose an appropriate set of fluents, but it would certainly be part of such an account that all fluents should correspond to projectable predicates, in Goodman's sense.

4.5. Nonmonotonic Treatments of Inertia and a Package of Problems

The idea behind nonmonotonic solutions to the Frame Problem is to treat inertia as a default; changes are assumed to occur only if there is some special reason for them to occur. In an action-centered account of change, this means that absence of change is inferred when an action is performed unless a reason for the change can be found in axioms for the action.

For explicitness, I will use Reiter's default logic to illustrate the formalization. Recall that in Reiter's theory, defaults are represented as rules, not formulas, so that they are not

[26] See [Schubert, 1990, Reiter, 1993].

subject to quantification. To formalize inertia, then, we need to use default rule schemata. For each fluent f, action a, and situation s, the set of these schemata will include an instance of the following schema:

$$\mathbf{IR}(\mathbf{f}, \mathbf{a}, \mathbf{s}) \quad \frac{\top \;:\; \mathit{Holds}(\mathrm{f}, \mathrm{s}) \leftrightarrow \mathit{Holds}(\mathrm{f}, \textsc{Result}(\mathrm{a}, \mathrm{s}))}{\mathit{Holds}(\mathrm{f}, \mathrm{s}) \leftrightarrow \mathit{Holds}(\mathrm{f}, \textsc{Result}(\mathrm{a}, \mathrm{s}))} \;.$$

This way of doing things makes any case in which a fluent changes truth value a *prima facie* anomaly. But it follows from Reiter's account of extensions that such defaults are overridden when they conflict with the monotonic theory of situation dynamics. So if, for instance, there is a monotonic causal axiom for the action `blacken` ensuring that blackening a block will make it black in the resulting situation, then the appropriate instance of **IR** will be inefficacious, and there will be no extension in which a white block remains white when it is blackened.

The Frame Problem somehow managed to capture the attention of a wide community—but if one is interested in understanding the complex problems that arise in generalizing formalisms like the Situation Calculus, while at the same time ensuring that they deliver plausible solutions to a wide variety of scenarios, it is more useful to consider a larger range of problems. For the AI community, the larger problems include the Frame Problem itself, the Qualification Problem, the Ramification Problem, generalizability along a number of important dimensions including incomplete information, concurrency (multiple agents), and continuous change, and finally a large assortment of specific challenges such as the scenarios mentioned later in this section.

The *Qualification Problem* arises generally in connection with the formalization of commonsense generalizations. Typically, these involve exceptions, and these exceptions—especially if one is willing to entertain far-fetched circumstances—can iterate endlessly. The same phenomenon, under a label like 'the problem of *ceteris paribus* generalizations', is familiar from analytic philosophy. It also comes up in the semantics of *generic* constructions found in natural languages.[27] In a sense, this problem is addressed at a general level by nonmonotonic logics, which—though they do not provide a way to enumerate exceptions—do allow common sense generalizations to be formulated as defaults, as well as enabling further qualifications to be added nondestructively. Ideally, then, the initial generalization can be stated as an axiom and qualifications can be added incrementally in the form of further axioms.

The Qualification Problem was raised in [McCarthy, 1986], where it was motivated chiefly by generalizations concerning the consequences of actions; McCarthy considers in some detail the generalization that turning the ignition key in an automobile will start the car. Much the same point, in fact, can be made about virtually any action, including stacking one block on another—the standard action that is used to illustrate the Situation Calculus. A circumscriptive approach to the Qualification Problem is presented in [Lifschitz, 1987]; this explicitly introduces the precondition relation between an action and its preconditions into the formalism, and circumscriptively minimizes preconditions, eliminating from preferred models any "unknown preconditions" that might render an action inefficacious.

Several dimensions of the Qualification Problem remain as broad, challenging research problems. For one thing, not every nonmonotonic logic provides graceful mechanisms for

[27] See [Carlson and Pelletier, 1995].

qualification. default logic, for instance, does not deliver the intuitively desired conclusions. Suppose one formalizes the common sense generalization that if I press the 'a' key on my computer it will type 'a' as a normal default:

$$\frac{\mathsf{T} \,:\, \textsc{Value}(\texttt{text}, \textsc{Result}(\texttt{Press-a}, \mathrm{s})) = \textsc{Value}(\texttt{text}, \mathrm{s}) + \text{`a'}}{\textsc{Value}(\texttt{text}, \textsc{Result}(\texttt{Press-a}, \mathrm{s})) = \textsc{Value}(\texttt{text}, \mathrm{s}) + \text{`a'}}\,.$$

If I then formalize the exception to this generalization that if I press the 'a' key while the Alt key is depressed the cursor moves to the beginning of the current sentence as a normal default along the same lines, I get *two* extensions: one in which pressing 'a' while the Alt key is depressed adds 'a' to the text and another in which it moves the cursor. The problem is that default logic does not provide for more specific defaults to override ones that are more general. This principle of specificity has been discussed at length in the literature. Incorporating it in a nonmonotonic logic can complicate the theory considerably; see, for instance, [Asher and Morreau, 1991] and [Horty, 1994]. And, as [Elkan, 1995] points out, the Qualification Problem raises computational issues.

Relatively little attention has been given to the Qualification Problem for characterizing actions, in comparison with other problems in temporal reasoning. In particular, the standard accounts of *unsuccessful* actions are somewhat unintuitive. In the formalization of [Lifschitz, 1987], for instance, actions with some unsatisfied preconditions are only distinguished from actions whose preconditions all succeed in that the conventional effects of the action will only be ensured when the preconditions are met. It is as if an action of spending $1,000,000 can be performed at any moment—although if you don't have the money, no effects in particular will be guaranteed.[28] And there is no distinction between actions that cannot even be attempted (like boarding a plane in London when you are in Sydney), actions that can be attempted, but in which the attempt can be expected to go wrong (like making a withdrawal when you have insufficient funds), actions that can be attempted with reasonable hope of success, and actions that can be attempted with guaranteed success.

As J.L. Austin made clear in ([Austin, 1961]), the ways in which actions can be attempted, and in which attempted actions can fail, are a well developed part of common sense reasoning. Obviously, in contemplating a plan containing actions that may fail, one may need to reason about the consequences of failure. Formalizing the pathology of actions, providing a systematic theory of ways in which actions and the plans that contain them can go wrong, would be a useful addition to planning formalisms, and one that would illuminate important themes in philosophy.

The challenge posed by the *Ramification Problem* (characterized first in [Finger, 1987]) is to formalize the indirect consequences of actions, where "indirect" effects are not delayed[29], but are temporally immediate but causally derivative. If I walk into a room, the direct effect is that I am in now the room. There are also many indirect effects: for instance, my shirt also is now in the room. You can see from this that the formulation of the problem

[28]This way of putting it is a little misleading for the Situation Calculus, since there is no robust notion of performing an action; instead, you consider the results of performing hypothetical action sequences. Even so, the point that the theory of unsuccessful actions has not been explored holds up.

[29]Effects of actions that are delayed in time are a separate problem, which, as far as I know, no one has solved.

presupposes a distinction between direct consequences of actions (ones that attach directly to an action, and that are ensured by the successful performance of the action) and other consequences. This assumption is generally accepted without question in the AI literature on action formalisms. You can make a good case for its common sense plausibility—for instance, many of our words for actions ('to warm', to 'lengthen', 'to ensure') are derived from the effects that are conventionally associated with them. And in these cases, success is entailed: if someone has warmed something, this entails that it became warm.[30] A typical example is discussed in [Lin, 1995]: a certain suitcase has two locks, and is open if and only if both locks are open. Then (assuming that actions are not performed concurrently) opening one lock will open the suitcase if and only if the other lock is open. Here, opening a lock is an action, with direct consequences; opening a suitcase is not an action, it is an indirect effect.

Obviously, the Ramification Problem is intimately connected with the Frame Problem. In approaches that adopt a nonmonotonic inertial axiom to solve the Frame Problem, inertial defaults will need to be overridden by conclusions about ramifications in order to obtain correct results. In case the left lock of the suitcase is open, for instance, and an action of opening the right lock is performed, then the default conclusion that the suitcase remains closed, simply because it is closed initially, needs somehow to be overridden. The most detailed and promising approaches to the Ramification Problem depend on the development of theories of common sense causation, and therefore are closely related to the causal approaches to reasoning about time and action discussed below in Section 4.6. See, for instance, [Lin, 1995, Thielscher, 1989, Giunchiglia *et al.*, 1997].

Philosophical logicians have been content to illustrate their ideas with relatively small-scale examples. The formalization of even large-scale mathematical theories is relatively unproblematic. Logicist AI is the first branch of logic to undertake the task of formalizing large examples involving nontrivial common sense reasoning. In doing so, the field has had to invent new methods. An important part of the methodology that has emerged in formalizing action and change is the prominence that is given to challenges, posed in the form of *scenarios*. These scenarios represent formalization problems which usually involve relatively simple, realistic examples designed to challenge the logical theories in specific ways. Typically, there will be clear common sense intuitions about the inferences that should be drawn in these cases. The challenge is to design a logical formalism that will provide general, well-motivated solutions to these benchmark problems.

Among the many scenarios that have been discussed in the literature are the Baby Scenario, the Bus Ride Scenario, the Chess Board Scenario, the Ferryboat Connection Scenario, the Furniture Assembly Scenario, the Hiding Turkey Scenario, the Kitchen Sink Scenario, the Russian Turkey Scenario, the Stanford Murder Mystery, the Stockholm Delivery Scenario, the Stolen Car Scenario, the Stuffy Room Scenario, the Ticketed Car Scenario, the Walking Turkey Scenario, and the Yale Shooting Anomaly. Accounts of these can be found in [Shanahan, 1997] and [Sandewall, 1994]; see especially [Sandewall, 1994][Chapters 2 and 7].

[30] The relationship between an action and the occurrence of its conventional consequences is complicated, of course, by the "imperfective paradox" (see [Dowty, 1977, Lascarides, 1992]). Some of the work on AI theories of action and change is informed by these complexities; see [Steedman, 1998, Steedman, 1995]. But for the most part, they have not been taken into account in the AI literature.

Many of these scenarios are designed to test advanced problems that I will not discuss here—for instance, challenges dealing with multiple agents, or with continuous changes. Here, I will concentrate on one of the earliest, and probably the most subtle of these scenarios: the Yale Shooting Anomaly, first reported in [Hanks and McDermott, 1985] and published in [Hanks and McDermott, 1986, Hanks and McDermott, 1987].

The Yale Shooting Anomaly involves three actions: `load`, `shoot`, and `wait`. A propositional fluent `Loaded` tracks whether a certain pistol is loaded; another fluent, `Alive`, tracks whether a certain person, Fred, is alive. `load` has no preconditions; its only effect is `Loaded`. `shoot` has `Loaded` as its only precondition and `Alive` as a negative effect; `wait` has no preconditions and no effects.

Causal information regarding the axioms is formalized as follows.

Load $\forall \mathrm{s}\, \mathit{Holds}(\texttt{load}, \textsc{Result}(\texttt{load}, \mathrm{s}))$

Shoot 1 $\forall \mathrm{s}[\mathit{Holds}(\texttt{Loaded}, \mathrm{s}) \rightarrow \mathit{Holds}(\neg\texttt{Alive}, \textsc{Result}(\texttt{shoot}, \mathrm{s}))]$

Shoot 2 $\forall \mathrm{s}[\mathit{Holds}(\texttt{Loaded}, \mathrm{s}) \rightarrow \mathit{Holds}(\neg\texttt{Loaded}, \textsc{Result}(\texttt{shoot}, \mathrm{s}))]$

There is no Wait Axiom—that is, `wait` has no preconditions and no effects.

We will formalize the inertial reasoning in this scenario using a nonmonotonic logic—to be specific, we use Reiter's default logic. The set D of defaults for this theory consists of all instances of the inertial schema **IR**.

In the initial situation, Fred is alive and the pistol is unloaded.

IC 1 $\mathit{Holds}(\texttt{Alive}, \mathrm{s}_0)$

IC 2 $\neg\mathit{Holds}(\texttt{Loaded}, \mathrm{s}_0)$

The monotonic theory W of the scenario consists of: (1) the action axioms **Load**, **Shoot 1** and **Shoot 2** and (2) the initial conditions **IC1** and **IC 2**.

Let $\mathrm{s}_1 = \textsc{Result}(\texttt{load}, \mathrm{s}_0)$, $\mathrm{s}_2 = \textsc{Result}(\texttt{wait}, \mathrm{s}_1)$, and $\mathrm{s}_3 = \textsc{Result}(\texttt{shoot}, \mathrm{s}_2)$.

The Yale Shooting Anomaly consists of the fact that the theory allows an extension in which the actions are `load`; `shoot`; `wait`, and in the final situation s_3, the pistol is unloaded and Fred is alive. The initial situation in the Anomaly and the three actions, with their resulting situations, can be pictured as follows.

$$\mathrm{s}_0 \xrightarrow{\texttt{load}} \mathrm{s}_1 \xrightarrow{\texttt{wait}} \mathrm{s}_2 \xrightarrow{\texttt{shoot}} \mathrm{s}_3$$

The natural, expected outcome of these axioms is that the pistol is loaded and Fred is alive after waiting, so that shooting yields a final outcome in which Fred is not alive and the pistol is unloaded. There is no problem in showing that this corresponds to an extension; the problem is the presence of the other, anomalous extension, which looks like this.

$$\begin{array}{ccccccc}
\texttt{Alive} & & \texttt{Alive} & & \texttt{Alive} & & \texttt{Alive} \\
\neg\texttt{Loaded} & & \texttt{Loaded} & & \neg\texttt{Loaded} & & \neg\texttt{Loaded} \\
\mathrm{s}_0 & \xrightarrow{\texttt{load}} & \mathrm{s}_1 & \xrightarrow{\texttt{wait}} & \mathrm{s}_2 & \xrightarrow{\texttt{shoot}} & \mathrm{s}_3
\end{array}$$

Here is a narrative version of this extension. At first, Fred is alive and the pistol is unloaded. After loading, the pistol is loaded and Fred remains alive. After waiting the pistol becomes unloaded and Fred remains alive. Shooting is then vacuous since the pistol is unloaded, so finally, after shooting, Fred remains alive and the pistol remains unloaded. The best way to see clearly that this is an extension is to work through the proof. Less formally, though, you can see that the expected extension violates just one default: the frame default for `Alive` is violated when Fred changes state in the last step. But the anomalous extension also violates only one default: the frame default for `Loaded` is violated when the pistol spontaneously becomes unloaded while waiting. So, if you just go by the number of defaults that are violated, both extensions are equally good.

The Yale Shooting Anomaly represents a major obstacle in developing a theory of predictive reasoning. A plausible, well-motivated logical solution to the Frame Problem runs afoul of a simple, crisp example in which it clearly delivers the wrong results. Naturally, the literature concerning the Yale Shooting Problem is extensive. Surveys of some of this work, with bibliographical references, can be found in [Morgenstern, 1996, Shanahan, 1997].

4.6. Some Emergent Frameworks

Many formalisms have been proposed to deal with the problems surveyed in the previous section. Some are more or less neglected today. Several are still advocated and defended by leading experts; some of these are associated with research groups who are not only interested in developments of logical theory, but in applications in planning and cognitive robotics.

The leading approaches provide solutions to the main problems mentioned in Section 4.5, and to many of the scenarios designed to test and illustrate theories of reasoning about action and change. It is commonly agreed that good solutions need to be generalizable to more complex cases than the early planning formalisms, and that in particular the solutions they offer should be deployable even when continuous time, concurrent actions, and various kinds of ignorance are allowed. Also, it is generally agreed that the formalisms should support several kinds of reasoning, and, in particular, not only prediction and plan verification but *retrodiction*, i.e., construction of a sequence of states and actions given partial information in the form of a narrative.

I will describe four approaches here: (1) Features and fluents (Sandewall), (2) Motivated Action Theory (Morgenstern and Stein), (3) State Minimization in the Event Calculus (Shanahan) and Causal Theories (Lifschitz and others). My accounts of the first three will be fairly brief; fortunately, each approach is well documented in a single reference. I believe that the fourth approach is likely to be most interesting to philosophers, and that it contains elements that will be of lasting importance whatever changes future developments in this area may bring.

Features and fluents This approach, described in [Sandewall, 1994], uses preference semantics as a way to organize nonmonotonic solutions to the problems of reasoning about action and change. Rather than introducing a single logical framework, Sandewall considers a number of temporal logics, including ones that use discrete, continuous, and branching time. The properties of the logics are systematically tested against a large suite of test scenarios.

Motivated Action Theory This theory grew out of direct consideration of the problems in temporal reasoning described above in Section 4.5, and especially the Yale Shooting Scenario. In [Morgenstern and Stein, 1994], Morgenstern and Stein seek to find a general, intuitively motivated logical framework that solves the difficulties. They settle on the idea that unmotivated actions are to be minimized, where an action ("actions" construed generally enough to include any change) can be motivated directly, e.g. by an axiom, or indirectly, through chains of motivations. The key technical idea of the paper is a (rather complicated) definition of motivation in an interval-based temporal logic. In [Morgenstern, 1996], Morgenstern presents a summary of the theory, along with reasons for rejecting its causal rivals. The most important of these reasons is that these theories, based on the Situation Calculus, do not appear to generalize to cases allowing for concurrency and ignorance. She also cites the failure of early causal theories to deal with retrodiction.

State-Based Minimization in the Event Calculus In [Baker, 1989], Andrew Baker presented a solution to the version of the Yale Shooting problem in the Situation Calculus, using a circumscriptive inertial axiom. The very brief account of circumscription above in Section 3 indicated that circumscription uses preferred models in which the extensions of certain predicates are minimized. In the course of this minimization, a set of parameters (including, of course, the predicates to be minimized) is allowed to vary; the rest are held constant. Which parameters vary and which are held constant is determined by the application.

In the earliest circumscriptive solutions to the Frame Problem, the inertial rule **CIR** is stated using an abnormality predicate.

$$\textbf{CIR}\ \ \forall \mathrm{f}\forall \mathrm{s}\forall \mathrm{a}[\neg Ab(\mathrm{f},\mathrm{a},\mathrm{s}) \rightarrow [\mathit{Holds}(\mathrm{f},\mathrm{s}) \leftrightarrow \mathit{Holds}(\mathrm{f},\textsc{Result}(\mathrm{a},\mathrm{s}))]]$$

This axiom uses a biconditional, so that it can be used for retrodiction; this is typical of the more recent formulations of common sense inertia.

In circumscribing, the abnormality predicate is minimized while the *Holds* predicate is allowed to vary and all other parameters are fixed. This formalization succumbs to the Yale Shooting Anomaly in much the same way that default logic does. (Circumscription does not involve multiple extensions, so the problem emerges as the nonderivability of the conclusion that Fred is alive after the occurrence of the shooting.)

In Baker's reformulation of the problem, separate axioms ensure the existence of a situation corresponding to each Boolean combination of fluents, and the RESULT function is allowed to vary, while the *Holds* predicate is held constant. In this setting, the RESULT function needs to be specified for "counterfactual" actions—in particular, for shooting as well as for waiting in the Yale Shooting Anomaly. It is this feature that eliminates the incorrect model for that scenario; for details, see [Baker, 1989] and [Shanahan, 1997, Chapter 6].

This idea, which Shanahan calls "State-Based Minimization," is developed and extended in [Shanahan, 1997], in the context of a temporal logic deriving from the Event Calculus of Kowalski and Sergot; see [Kowalski and Sergot, 1986]. Shanahan's formalism has the advantage of being closely connected to implementations using logic programming.

Causal Theories Recall that in the anomalous model of the Yale Shooting Scenario the gun becomes unloaded after the performance of the `wait` action, an action which has no conventional effects—the unloading, then, is uncaused. In the context of a nonmonotonic logic—and without such a logic, the Yale Shooting Anomaly would not arise—it is very natural to formalize this by treating uncaused eventualities as abnormalities to be minimized.

This strategy was pursued by Hector Geffner in [Geffner, 1990, Geffner, 1992], where he formalizes this simple causal solution to the Yale Shooting Anomaly. But the solution is presented in the context of an ambitious general project in nonmonotonic logic that not only develops properties of the preferred model approach and shows how to apply it to a number of reasoning problems, but that relates nonmonotonic logic to probabilities, using ideas deriving from [Adams, 1975]. In [Geffner, 1992], the causal theory is sketched; it is not developed to show its adequacy in dealing with the battery of problems presented above, and in particular the Ramification Problem is left untouched.

The work beginning with [Lifschitz, 1987] has contributed to a sustained line of research in the causal approach—not only by Lifschitz and students of his such as Enrico Giunchiglia and Hudson Turner, but by researchers at other sites. For work in this area, and further references, see [Haugh, 1987, Elkan, 1991, Baral, 1995, Lin, 1995, McCain and Turner, 1995, Gustaffson and Doherty, 1996, Thielscher, 1996, McCain and Turner, 1997, Thielscher, 1989, Gelfond and Lifschitz, 1998, Lifschitz, 1997, Nakashima *et al.*, 1997, Giunchiglia and Lifschitz, 1998, Lifschitz, 1998, Turner, 1999].

Here, I will briefly describe some of the features of the theory presented in [Turner, 1999], which returns to the ideas of [Geffner, 1992], but places them in a simpler logical setting and applies them to the formalization of more complex scenarios that illustrate the interactions of causal inertia with other considerations, especially the Ramification Problem.

The idea is to treat `Caused` as a modal operator $[c]$, making this the basis of a modal nonmonotonic logic. In the preferred models of this logic, the caused propositions coincide with the propositions that are true, and this must be the only possibility consistent with the extensional part of the model. To make this more explicit, recall that in the possible worlds interpretation of **S5**, it is possible to identify possible worlds with *state descriptions*, which we can represent as sets I of literals (atomic formulas and their negations). Making this identification, then, we can think of a model as a pair $\langle I, S\rangle$, where S is a set of interpretations including I. The modal operator $[c]$ is given the standard semantics: let S be a set of interpretations. Then, where $I \in S$, $S \models_I [c]A$ if and only if $S \models_{I'} A$ for all $I' \in S$. $\langle I, S\rangle$ *satisfies* a set of formulas T if and only if $S \models_I A$ for all $A \in T$.

Turner's preferred models of T are the pairs $\langle I, S\rangle$ such that: (1) $\langle I, S\rangle$ satisfies T, (2) $S = \{I\}$, and (3) $\langle I, S\rangle$ is the unique interpretation $\langle I', S'\rangle$ meeting conditions (1) and (2) with $I' = I$. Condition (2) guarantees the "universality of causation"; it validates $A \leftrightarrow [c]A$. Condition (3) "grounds" causality in noncausal information (in the models in which we are interested, this will be information about the occurrence of events), in the strongest sense: it is uniquely determined by this information.

Although it is not evident from the formulation, Turner's account of preferred models is related to the constructions of more general nonmonotonic logics, such as default logic. Consult [Turner, 1999] for details.

The axioms that specify the effects of actions treat these effects as caused; for instance, the axiom schema for loading would read as follows:

Causal-Load $\texttt{[}c\texttt{]}\mathit{Holds}(\texttt{load}, \textsc{Result}(\texttt{load}, \mathrm{s})$[31]

Ramifications of the immediate effects of actions are also treated as caused. And the nonmonotonic inertial axiom schemata take the form

$$[[\texttt{[}c\texttt{]}\mathit{Holds}(\mathrm{f}, \mathrm{s})] \wedge \mathit{Holds}(\mathrm{f}, \textsc{Result}(a, \mathrm{s}))] \rightarrow \texttt{[}c\texttt{]}\mathit{Holds}(\mathrm{f}, \textsc{Result}(a, \mathrm{s}))$$

and

$$[[\texttt{[}c\texttt{]}\neg\mathit{Holds}(\mathrm{f}, \mathrm{s})] \wedge \neg\mathit{Holds}(\mathrm{f}, \textsc{Result}(a, \mathrm{s}))] \rightarrow \texttt{[}c\texttt{]}\neg\mathit{Holds}(\mathrm{f}, \textsc{Result}(a, \mathrm{s})).$$

Thus, a true proposition can be caused either because it is the direct or indirect effect of an action, or because it involves the persistence of a caused proposition. Initial conditions are also considered to be caused, by stipulation.

To illustrate the workings of this approach, let's consider the simplest case of inertia: we have a language with just one constant denoting a fluent, f, and one action-denoting constant, `wait`. As in the Yale Shooting Problem, there are no axioms for `wait`; this action can always be performed and has no associated effects. Let s_1 be $\textsc{Result}(\texttt{wait}, \mathrm{s}_0)$. The theory T contains an initial condition for f, $\mathit{Holds}(\mathrm{f}, \mathrm{s}_0)$ and a statement that the initial condition is caused, $\texttt{[}c\texttt{]}\mathit{Holds}(\mathrm{f}, \mathrm{s}_0)$, as well as the inertial schemata.

Two models of T satisfy conditions (1) and (2): $M_1 = \langle I_1, \{I_1\}\rangle$ and $M_2 = \langle I_2, \{I_2\}\rangle$, where $I_1 = \{\mathit{Holds}(\mathrm{f}, \mathrm{s}_0), \mathit{Holds}(\mathrm{f}, \mathrm{s}_1)\}$ and $I_2 = \{\mathit{Holds}(\mathrm{f}, \mathrm{s}_0), \neg\mathit{Holds}(\mathrm{f}, \mathrm{s}_1)\}$.

M_1 is the intended model, in which nothing changes. It satisfies Condition (3), since if $\langle I_1, S\rangle$ satisfies T it satisfies $\texttt{[}c\texttt{]}\mathit{Holds}(\mathrm{f}, \mathrm{s}_1)$ by the inertial axiom

$$[\texttt{[}c\texttt{]}\mathit{Holds}(\mathrm{f}, \mathrm{s})] \wedge \mathit{Holds}(\mathrm{f}, \mathrm{s}_1))] \rightarrow \texttt{[}c\texttt{]}\mathit{Holds}(\mathrm{f}, \mathrm{s}_1).$$

Therefore, $S = \{I_1\}$.

M_2 is an anomalous model, in which the fluent ceases spontaneously. This model does not satisfy Condition (3), since $M_3 = \langle I_2, \{I_1, I_2\}\rangle$ also satisfies T; in particular, it satisfies the inertial axiom for f because it fails to satisfy $\mathit{Holds}(\mathrm{f}, \mathrm{s}_1)$. So, while M_1 is a preferred model, M_2 is not.

The apparent usefulness of a "principle of universal causality" in accounting for a range of problems in qualitative common sense reasoning will be tantalizing to philosophers. And the causal theory, as initiated by Geffner and developed by Turner, has many interesting detailed features. For instance, while philosophical work on causality has concentrated on the causal relation, this work in logical AI shows that a great deal can be done by using only a nonrelational causal predicate.

Morgenstern's two chief criticisms of the causal approach to reasoning about actions are that it does not give an adequate account of explanation[32] and that the logical context in which it works (the Situation Calculus) is limited. As work on the approach continues, progress is being made in these areas. But the constraints that a successful logic of action and change must meet are so complex that it is a reasonable research methodology to concentrate initially on a restricted logical setting.

[31] Turner uses a discrete temporal logic other than the Situation Calculus. But for uniformity of presentation I have used the Situation Calculus to present the ideas.

[32] In explanation problems, one is reasoning backwards in time. Here, information is provided about a series of occurring states and the problem is to provide actions that account for the occurrences.

4.7. Action Formalisms and Natural Language

Although for many AI logicists, the goal of action formalisms is to illuminate an important aspect of common sense reasoning, most of their research is uninformed by an important source of insights into the common sense view of time—namely, natural language. Linguists concerned with the semantics of temporal constructions in natural language, like the AI community, have begun with ideas from philosophical logic but have discovered that these ideas need to be modified in order to deal with the phenomena. A chief discovery of the AI logicists has been the importance of actions and their relation to change. Similarly, an important discovery of the "natural language logicists" has been the importance of different kinds of events (including structured composite events) in interpreting natural language. From work such as this the idea of "natural language metaphysics" (see, for instance, [Bach, 1989]) has emerged.

The goal of articulating a logical framework tailored to a representational system that is motivated by systematic evidence about meanings in natural languages is not acknowledged by all linguistic semanticists. Nevertheless, it is a significant theme in the linguistic literature. This goal is remarkably similar to those of the common sense logicists, but the research methodology is entirely different.

Can the insights of these separate traditions be reconciled and unified? Is it possible to constrain theories of temporal representations and reasoning with the insights and research methodologies of both traditions? In [Steedman, 1995, Steedman, 1998] these important questions are addressed, and a theory is developed that extends action formalisms like the Situation Calculus, and that incorporates many of the insights from linguistic semantics. The project reported in [Steedman, 1998] is still incomplete, but the results reported there make a convincing case that the event-based ideas from linguistics can be fruitfully combined with the action-centered formalisms in the AI literature. The possibility of this unification is one of the most exciting logical developments in this area, bringing together as it does two independent descendants of the earlier work in the logic of time.

5. Causal reasoning

In Section 4.6, we traced the reasons for the development of theories incorporating causality in work on reasoning about action and change. This is not the only area of AI in which causality has emerged. Causality figures in qualitative reasoning about devices; for Herbert Simon's important work in this area, which goes back to the 1950s, see [Simon, 1952, Simon, 1977, Iwasaki and Simon, 1986]. Both these traditions are important. But the most robust and highly developed program in AI relating to causality is that of Judea Pearl and his students and associates, which derives from the use of causal diagrams in the formalism for reasoning about probabilities known as *Bayesian Belief Networks*.

Pearl's program has developed into a far-reaching campaign to rehabilitate causality in statistical thinking. I will not discuss this topic here. For one thing, this survey omits probabilistic reasoning in AI. For another, Pearl's views on causality are systematically and comprehensively presented in a recent book-length study, [Pearl, 2000].

But I do wish to point out that the work on causality discussed in Section 4.6 and Pearl's ideas do share some common themes. On both approaches: *action* is central for causality.

Also there is a focus on causality as a tool in reasoning that is necessitated in part by limited resources. Another important theme is the deployment and systematic study of formalisms in which causality is related to other constructs (in particular, to probability and to qualitative change) and a variety of realistic reasoning problems are addressed.

These commonalities provide reason to hope that we will see a science of causality emerging from the AI research, unifying the contributions of the probabilistic, the qualitative physics, and the nonmonotonic traditions, and illuminating the various phases of causal reasoning.

Whether you take causality to be a fundamental construct in natural science, or a fundamental natural phenomenon, depends on whether you have in mind an idealized nature described by differential equations or you have in mind the view of nature we have to take in order to act, either in everyday situations, or for that matter in designing experiments in the laboratory. The fact that, as Bertrand Russell noted (see [Russell, 1957]), causality is not to be found as a theoretical primitive in contemporary physical theories is at odds with its seeming importance in so many familiar areas of reasoning. The rigorous theories emerging in AI that are beginning to illuminate the workings of causality are important not only in themselves, but in their potentiality to illuminate wider philosophical issues.

6. Spatial reasoning

The precomputational literature in philosophical logic relating to spatial reasoning is very sparse in relation, for instance, to the temporal literature. The need to support computational reasoning about space, however, in application areas such as motion planning and manipulation in physical space, the indexing and retrieval of images, geographic information systems, diagrammatic reasoning, and the design of high-level graphics programs has led to new interest in spatial representations and spatial reasoning. Of course, the geometrical tradition provides an exceptionally strong mathematical resource for this enterprise. But as in many other AI-related areas, it is not clear that the available mathematical theories are appropriate for informing these applications, and many computer scientists have felt it worthwhile to develop new foundations. Some of this work is closely related to the research in qualitative reasoning mentioned above in Section 2.2, and in some cases has been carried out by the same individuals.

The literature in spatial reasoning is extensive; for references to some areas not discussed here, see [Forbus *et al.*, 1991, Kapur and Mundy, 1988, Renz and Nebel, 1999, Stock, 1997, Wilson, 1998, Yeap and Jeffries, 1999, Allwein and Barwise, 1996, Glasgow *et al.*, 1995, Hammer, 1995, Kosslyn, 1990, Osherson and Lasnik, 1990, Burger and Bhanu, 1992, Chen, 1990]. Here, I will discuss only one trend, which is closely connected with parallel work in philosophical logic.

Qualitative approaches to space were introduced into the logical literature early in the twentieth century by Lesniewski; see [Leśniewski, 1916], which presents the idea of a *mereology*, or qualitative theory of the part-whole relation between physical individuals. This idea of a logical theory of relations among regions or the objects that occupy them, which does not depend on construing regions as sets of points, remained an active area of philosophical logic, even though it attracted relatively few researchers. More recent work in the philosophical literature, especially [Clarke, 1981, Clarke, 1985, Simons, 1987, Casati and Varzi, 1996, Casati and Varzi, 1999], was directly influential on current computational work.

The Regional Connection Calculus (RCC), developed by computer scientists at the University of Leeds, is based on a primitive C relating regions of space: the intended interpretation of $C(x, y)$ is that the intersection of the closures of the values of x and y is nonempty. (See [Cohn *et al.*, 1997, Cohn, 1996] for details and references.) One area of research concerns the definability of shapes in RCC. The extent of what can be defined with this simple primitive is surprising, but the technicalities quickly become complex; see, for instance, [Gotts, 1994, Gotts, 1996]. The work cited in [Cohn *et al.*, 1997] describes constraint propagation techniques and encodings in intuitionistic propositional logic as ways of supporting implemented reasoning based on RCC and some of its extensions. More recent work based on RCC addresses representation and reasoning about motion, which of course combines spatial and temporal issues; see [Wolter and Zakharyaschev, 2000]. For more information about qualitative theories of movement, with references to other approaches, see [Galton, 1997].

7. Reasoning about knowledge

Epistemic logic is another area in which strong influences from philosophical logic can be traced on logic in computer science. The classical source for epistemic logic is [Hintikka, 1962], in which Jaakko Hintikka showed that a modal approach to single-agent epistemic attitudes could be informative and rewarding. This work discusses at length the question of exactly which constraints are appropriate for knowledge and belief, when these attitudes are viewed as explicated by a model theoretic relation over possible worlds; in both cases, Hintikka argues for **S4** type operators.

In several papers (including [McCarthy, 1979]), John McCarthy has recommended an approach to formalizing knowledge that uses first-order logic, but that quantifies explicitly over such things as individual concepts. In this section I'll discuss the approach taken by most computer scientists, however, who use a modal language to formalize propositional attitudes.

The logical aspects of modal epistemic logic were not significantly developed after Hintikka's 1962 presentation; instead, the philosophical literature (which is not extensive, compared with many other topics in the area) concentrates on the issue of *hyperintensionality,* or closure of epistemic attitudes under logical consequence. This topic is especially challenging, turning out to be closely related to the semantic paradoxes, and the philosophical literature is inconclusive. Intuitions seem to conflict, and it is difficult to find ways to model the important phenomena using logical techniques.[33]

[Fagin *et al.*, 1984] begins a tradition in computational logic that revives the modal approach to epistemic logic, developing generalized logical foundations and applications that had not occurred to the philosophers. The technical idea is to simplify the modality, using **S5** (or deontic **S5** for belief), but to introduce multiple agents, and to concentrate on reasoning having to do with agents' attitudes about one another's attitudes. Such logics have direct applications in the analysis of *distributed systems*, dynamic systems in which change is effected by message actions, which change the knowledge of agents according to rules determined by a *communications protocol.*

[33]For information about the philosophical tradition, see [Hintikka, 1986]. Also, see [Laux and Wansing, 1995].

As such, this work belongs to a separate area of computer science, but one that overlaps to some extent with AI. Later, this work has interacted with a research tradition in economics that is concerned with the role of knowledge in games and bargaining; see, for instance, [Geanakopolos, 1994, Osborne and Rubenstein, 1994][Chapter 5].

For some reason, the multi-agent case did not occur to philosophical logicians.[34] This is another example of the way in which need for an application (in this case, the need for a theory of distributed systems) provided the inspiration for an important logical development. I will not present details concerning the logic here, since they are extensively and systematically recorded in [Fagin *et al.*, 1995]; this is essential reading for anyone seriously interested in this topic.

Much of the interdisciplinary work in applications of the logic of knowledge is reported in the proceedings of a series of conferences initiated in 1986 with [Halpern, 1986]. These conferences record one of the most successful collaborations of philosophers with logicians in Computer Science, although the group of involved philosophers has been relatively small. The focus of the conferences has gradually shifted from Computer Science to Economics.

AI applications deal with with knowledge in the form of stored representations, and the tradition in AI with which we are concerned here thinks of reasoning as the manipulation of symbolic representations. Also, it is mainly due to AI that the problem of limited rationality has become a topic of serious interest, providing a counterbalance to the idealizations of philosophy and economics.[35] So you would think that a logical model of propositional attitudes that is committed to closure under logical consequence would be highly unpopular in AI. But this is not so; the possible worlds approach to attitudes is not only the leading theory in the areas discussed in [Fagin *et al.*, 1995], but has even been advocated in robotics applications; see [Rosenschein, 1989, Rosenschein and Kaelbling, 1995].

Nevertheless, the issue of hyperintensionality has been investigated in the AI literature; see [Konolige, 1986, Lakemeyer, 1997, Levesque, 1984, Perlis, 1985]. Though there are some new positive results here, the AI work in this area, in my opinion, has been as inconclusive as that in philosophy.

The philosophical literature on a related topic, the logic of perception, has not been extensive; the main reference is [Hintikka, 1970].[36] But sensation is addressed in recent work in the AI literature which is concerned with developing logical frameworks for general-purpose applications in Robotics. The main idea in this area is to add sensing actions to the repertoire of a planning formalism of the sort discussed in Section 4. The earliest work in this area was carried out in the 1980s by Robert Moore; see [Moore, 1985, Moore, 1995b]. For some of the contemporary work in Cognitive Robotics, see [Bacchus *et al.*, 1999, Bacchus *et al.*, 1999, Baral *et al.*, 2000, Golden and Weld, 1996, Pirri and Finzi, 1999, Thielscher, 2000].

[34] A personal recollection: I was certainly aware of this case in the early 1970s, but did not devote much attention to it because it seemed to me that the generalization from the single-agent case was relatively trivial and did not pose any very interesting logical challenges.

[35] See, for instance, [Simon, 1982a, Simon, 1982b, Russell and Wefald, 1991].

[36] Although this topic has received attention more recently in Situation Theory, the logical issues, in my opinion, have not been illuminated by this work.

8. Logical approaches to natural language and communication

Over the last twenty-five years or so, many profound relations have emerged between logic and grammar. Computational linguistics (or natural language processing) is a branch of AI, and it is fairly natural to classify some of these developments under logic and AI. But many of them also belong to an independent tradition in logical foundations of linguistics; and in many cases it is hard (and pointless) to attempt a classification. This sketch will concentrate on developments that have to do with reasoning about linguistics; other applications of logic to linguistics are described in [van Benthem and ter Meulen, 1996].

9. Parsing and deduction

Grammar formalisms—special-purpose systems for the description of linguistic systems and subsystems—can be thought of as logics designed to axiomatize the association of linguistic structures with strings of symbols. You might be able to infer from such a system, for instance, that 'assignments' is the plural form of the nominalization of the verb 'assign'. So you can look at the process of *parsing* a string of words—of finding the linguistic structures, if any, that are associated with it—as a search for a proof in a certain logical system.

This approach has been highly successful as an analytic tool. It makes model-theoretic techniques applicable to linguistic reasoning, This makes the underlying reasoning problems much more transparent, and makes it possible to apply many well-developed areas of logic to grammar formalisms. For more information on these topics, see [Shieber, 1992, Buszkowski, 1996].

10. Feature structure logic

The usefulness and scope of logical techniques in relation to linguistics is greatly increased by the development of techniques for analyzing the way information is attached to linguistic units. It is very natural to represent the information attaching, say, to a lexical item in the form of a set of functions (or attributes) that produce values in some linguistic domain. A pronoun x may have a number, a person, and a case: if $x =$ 'we' then

$$\begin{aligned}&\mathit{number}(x) = \text{plural},\\ &\mathit{person}(x) = \text{first},\\ &\mathit{case}(x) = \text{nominative}.\end{aligned}$$

In more general cases, the values of these functions may themselves be linguistic units that take on values for certain attributes.

Allowing these functions to be partial provides a useful informational representation of the stages of a linguistic parse; much of the work of parsing involves completing this partial information, subject to constraints imposed by linguistic agreement conditions. Feature structures have a natural algebraic treatment, and there is an elegant treatment of their logic. For more information and references, see [Rounds, 1996].

11. Logic and discourse

The reasoning associated with discourse is the probably the least well understood area of computational linguistics. Although logical techniques do not yet play a major role in discourse, they seem to offer one of the most promising ways of providing a uniform account of the many forms of reasoning that are involved in generating and interpreting language in interactive conversation.

I will briefly mention three contributions to this area. Building on the fact that the rules governing conversation are exception-ridden, Alex Lascarides and Nicholas Asher have developed techniques for formalizing discourse phenomena based on nonmonotonic logic; see [Asher and Lascarides, 1994, Asher and Lascarides, 1997]. Jerry Hobbs and various co-workers look at the inference processes used in discourse as *abductive*, and propose to formalize abduction as a search for a proof in which certain "low-cost" assumptions may be made which serve as data or additional axioms for the proof. [Hobbs *et al.*, 1993] shows how an impressive range of discourse phenomena can be formalized using this idea. In practice, this abductive account looks rather similar to that of Lascarides and Asher, because it involves deploying axioms about discourse (in the form of Horn clause rules supplemented with weights giving the assumption costs of premises) that in effect are nonmonotonic.

In more recent work, Matthew Stone shows in [Stone, 1998] how modal logic can inform the complex reasoning involved in natural language generation. Generating a coherent, appropriately phrased text that usefully performs a task-oriented communication task is difficult to formalize because it requires the integration of complex and sophisticated domain information with discourse planning, user modeling, and linguistic constraints. Stone shows that modal logic can be used to modularize the formalization of the information required in this task; he also shows how modal theorem proving can be used to implement the reasoning.

12. Taxonomic Representation and Reasoning

12.1. Concept-Based Classification

Traditionally, the task of representing large amounts of domain information for general-purpose reasoning has been one of the most important areas of knowledge representation. Systems that exploit the intuitive taxonomic organization of domains are useful for this purpose; taxonomic hierarchies not only help to organize the process of knowledge acquisition, but provide a useful connection to rule-based reasoning.[37]

For domains in which complex definitions are a natural way to organize information, knowledge engineering services based on definitions of concepts have been extremely successful. Like variable-free versions of first-order logic (see, for instance, [Quine, 1960]), these systems are centered on concepts or first-order predicates, and provide a number of mechanisms for their definition. The fundamental algorithm associated with these *taxonomic logics* is a classifier which inputs a system of definitions and outputs the entailment relations between defined and primitive concepts. For background on these systems, see [Brachman *et al.*, 1991, Woods and Schmolze, 1992].

The simplest taxonomic logics can be regarded as subsystems of first-order logic with

[37]See [Stefik, 1995] for background on considerations having to do with knowledge engineering.

complex predicates; but they have been extended in many ways, and the issues raised by many of these extensions overlap in many cases with topics in philosophical logic.

12.2. Nonmonotonic Inheritance

Much more complex logical issues arise when the organization of a domain into hierarchies is allowed to have exceptions. One way to approach this topic is to explore how to make a taxonomic logic nonmonotonic in its own right; but *nonmonotonic inheritance* is a topic in its own right. Although there are strong affinities to nonmonotonic logic, nonmonotonic logic relies more heavily on graph-based representations than on traditional logical ideas, and seems to provide a much finer-grained approach to nonmonotonic reasoning that raises entirely new issues, and which quickly becomes problematic. For this reason, systems of nonmonotonic inheritance tend to be expressively weak, and their relations to the more powerful nonmonotonic logic has never been fully clarified. For background on this topic, see [Thomason, 1992, Horty, 1994].

13. Contextual Reasoning

In the tradition in philosophical logic dealing with contextual effects on the interpretation of expressions, as well as in the more recent tradition in dynamic logic, context is primarily formalized as an assignment of values to variables, and the language is designed to make explicit reasoning about context either very limited or outright impossible.

Concern in AI about the representation of large and apparently heterogeneous domains and about the integration of disparate knowledge sources, as well as interests in formalizing common sense of the sort discussed in Section 2.2, above, have led to interest in the AI community in formalizing languages that take context into account more explicitly.

In [McCarthy, 1993b], McCarthy recommends the study of languages containing a construct

$$ist(c, \phi),$$

where *ist* is read "is-true." This is analogous to the *Holds* construct of the situation calculus—but now c stands for a context, and ϕ is a possibly complex propositional representation, which many (including McCarthy) take to refer to a sentence.

There are analogies here both to modal logic and to languages with an explicit truth predicate. But the applications that are envisioned for a logic of context create opportunities and problems that are in many ways new. For more about the logic of context, see [Guha, 1991, McCarthy and Buvač, 1998], and the papers in [Bouquet *et al.*, 1999, Akman *et al.*, 2001].

14. Prospects for a Logical Theory of Practical Reason

I believe there is reason to hope that the combination of logical methods with planning applications in AI can enable the development of a far more comprehensive and adequate theory of practical reasoning than has heretofore been possible. As with many problems having to do with commonsense reasoning, the scale and complexity of the formalizations that are required are beyond the traditional techniques of philosophical logic. However, computational methods of implementing and testing the formalizations and with areas such as cognitive robotics to serve as laboratories for developing and testing ideas, we can hope to radically advance a problem that has seen little progress since it was first proposed by Aristotle: the problem of devising a formalization of practical reasoning that is genuinely applicable to common sense reasoning problems.

The classical work in deontic logic that was begun by von Wright (see [von Wright, 1983]) is one source of ideas; see [van der Torre, 1997, Horty, 2001]. In fact, as the more recent work in deontic logic shows, nonmonotonic logic provides a natural and useful way to modify the classical deontic logic.

An even more robust account of practical reasoning begins to emerge when these ideas are supplemented with work on the foundations of planning and reasoning about action that were discussed in Section 4, above. But this development can be pursued even further, by extending the formalism to include preferences and intentions.[38]

Ultimately, what is needed is a model of an intelligent reasoning and acting agent. Developing such a model need not be entirely a matter of logic, but according to one school of thought, logic has a central role to play in it; see, for instance, [Baral and Gelfond, 2000, Burkhard *et al.*, 1998, Rao and Georgeff, 1991, Wobcke *et al.*, 1998].

15. Acknowledgments

I am grateful to John McCarthy, who read a draft of this article and provided extensive and helpful comments.

16. Bibliography

[Adams, 1975] Ernest W. Adams. *The Logic of Conditionals.* D. Reidel Publishing Co., Dordrecht, 1975.

[Aiello *et al.*, 1996] Luigia Carlucci Aiello, Jon Doyle, and Stuart Shapiro, editors. *KR'96: Principles of Knowledge Representation and Reasoning.* Morgan Kaufmann, San Francisco, 1996.

[Akman *et al.*, 2001] Varol Akman, Paolo Bouquet, Richmond Thomason, and Roger A. Young, editors. *Modeling and Using Context.* Springer-Verlag, Berlin, 2001.

[Alcourrón, 1995] Carlos E. Alcourrón. Defeasible logics: Demarcation and affinities. In Gabriella Crocco, Luis Fari nas del Cerro, and A. Herzig, editors, *Conditionals: From Philosophy to Computer Science*, pages 67–102. Oxford University Press, Oxford, 1995.

[Allen *et al.*, 1991] James F. Allen, Richard Fikes, and Erik Sandewall, editors. *KR'91: Principles of Knowledge Representation and Reasoning.* Morgan Kaufmann, San Mateo, California, 1991.

[38] For background on quantitative models of preference and decision, see [Doyle and Thomason, 1999]. For work in AI on intentions, see, for instance [Cohen and Levesque, 1990, Konolige and Pollack, 1993, Pollack, 1992, Sadek, 1992].

[Allwein and Barwise, 1996] Gerard Allwein and Jon Barwise, editors. *Logical Reasoning With Diagrams*. Oxford University Press, Oxford, 1996.

[Amarel, 1968] Saul Amarel. On representations of problems of reasoning about actions. In D. Mitchie, editor, *Machine Intelligence 3*, pages 131–171. Ellis Horwood, Chichester, England, 1968.

[Antoniou, 1997] Grigoris Antoniou. *Nonmonotonic Reasoning*. The MIT Press, Cambridge, Massachusetts, 1997.

[Arlo-Costa and Shapiro, 1992] Horacio Arlo-Costa and Scott Shapiro. Maps between nonmonotonic logic and conditional logic. In Bernhard Nebel, Charles Rich, and William Swartout, editors, *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, pages 553–564. Morgan Kaufmann, San Mateo, California, 1992.

[Asher and Lascarides, 1994] Nicholas Asher and Alex Lascarides. Intentions and information in discourse. In James Pustejovsky, editor, *Proceedings of the Thirty-Second Meeting of the Association for Computational Linguistics*, pages 35–41, San Francisco, 1994. Association for Computational Linguistics, Morgan Kaufmann.

[Asher and Lascarides, 1997] Nicholas Asher and Alex Lascarides. Lexical disambiguation in a discourse context. In James Pustejovsky and Brian Boguraev, editors, *Lexical Semantics: The Problem of Polysemy*, pages 69–108. Oxford University Press, Oxford, 1997.

[Asher and Morreau, 1991] Nicholas Asher and Michael Morreau. Commonsense entailment: a modal theory of nonmonotonic reasoning. In J. Mylopoulos and R. Reiter, editors, *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, pages 387–392, Los Altos, California, 1991. Morgan Kaufmann.

[Asher, 1995] Nicholas Asher. Commonsense entailment: A conditional logic for some generics. In Gabriella Crocco, Luis Fari nas del Cerro, and A. Herzig, editors, *Conditionals: From Philosophy to Computer Science*, pages 103–145. Oxford University Press, Oxford, 1995.

[Austin, 1961] John L. Austin. A plea for excuses. In J.O. Urmson and G.J. Warnock, editors, *Philosophical Papers*. Oxford University Press, Oxford, 1961.

[Bacchus *et al.*, 1999] Fahiem Bacchus, Joseph Y. Halpern, and Hector J. Levesque. Reasoning about noisy sensors and effectors in the situation calculus. *Artificial Intelligence*, 111(1–2):171–208, 1999.

[Bach, 1989] Emmon Bach. *Informal Lectures on Formal Semantics*. State University of New York Press, Albany, NY, 1989.

[Baker, 1989] Andrew B. Baker. A simple solution to the Yale shooting problem. In Ronald J. Brachman, Hector J. Levesque, and Raymond Reiter, editors, *KR'89: Principles of Knowledge Representation and Reasoning*, pages 11–20. Morgan Kaufmann, San Mateo, California, 1989.

[Baral and Gelfond, 2000] Chitta Baral and Michael Gelfond. Reasoning agents in dynamic domains. In Jack Minker, editor, *Logic-Based Artificial Intelligence*, pages 257–279. Kluwer Academic Publishers, Dordrecht, 2000.

[Baral *et al.*, 2000] Chitta Baral, Sheila McIlraith, and Tran Cao San. Formulating diagnostic reasoning using an action language with narratives and sensing. In Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, editors, *KR2000: Principles of Knowledge Representation and Reasoning*, pages 311–322, San Francisco, 2000. Morgan Kaufmann.

[Baral, 1995] Chitta Baral. Reasoning about actions: Non-deterministic effects, constraints, and qualification. In Chris Mellish, editor, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 2017–2023, San Francisco, 1995. Morgan Kaufmann.

[Belnap, Jr., 1996] Nuel D. Belnap, Jr. Agents in branching time. In Jack Copeland, editor, *Logic and Reality: Essays on the Legacy of Arthur Prior*, pages 239–271. Oxford University Press, Oxford, 1996.

[Benferat *et al.*, 1997] Salem Benferat, Didier Dubois, and Henri Prade. Nonmonotonic reasoning, conditional objects, and possibility theory. *Artificial Intelligence*, 92(1–2):259–276, 1997.

[Besnard, 1992] Philippe Besnard. *Default Logic*. Springer-Verlag, Berlin, 1992.

[Birnbaum, 1991] Lawrence Birnbaum. Rigor mortis: A response to Nilsson's 'Logic and artificial intelligence'. *Artificial Intelligence*, 47(1–3):57–77, 1991.

[Boolos, 1993] George Boolos. *The Logic of Provability*. Cambridge Universoti Press, Cambridge, England, 1993.

[Bouquet *et al.*, 1999] Paolo Bouquet, Luigi Serafini, Patrick Brézillon, Massimo Benerecetti, and Francesca Castellani, editors. *Modeling and Using Contexts: Proceedings of the Second International and Interdisciplinary Conference, CONTEXT'99*. Springer-Verlag, Berlin, 1999.

[Boutilier *et al.*, 1996] Craig Boutilier, Thomas Dean, and Steve Hanks. Planning under uncertainty: Structural assumptions and computational leverage. In Malik Ghallab and Alfredo Milani, editors, *New Directions in AI Planning*, pages 157–171. IOS Press, Amsterdam, 1996.

[Boutilier, 1992] Craig Boutilier. Conditional logics for default reasoning and belief revision. Technical Report KRR–TR–92–1, Computer Science Department, University of Toronto, Toronto, Ontario, 1992.

[Boutilier, 1996] Craig Boutilier. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic*, 25(3):263–305, 1996.

[Brachman *et al.*, 1989] Ronald J. Brachman, Hector J. Levesque, and Raymond Reiter, editors. *KR'89: Principles of Knowledge Representation and Reasoning.* Morgan Kaufmann, San Mateo, California, 1989.

[Brachman *et al.*, 1991] Ronald J. Brachman, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lori A. Resnik. Living with Classic: When and how to use a Kl-One-like language. In John F. Sowa, editor, *Principles of Semantic Networks*, pages 401–456. Morgan Kaufmann, San Mateo, California, 1991.

[Brewka *et al.*, 1997] Gerhard Brewka, Jürgen Dix, and Kurt Konolige. *Nonmonotonic Reasoning: An Overview.* CSLI Publications, Stanford, 1997.

[Brewka, 1991] Gerhard Brewka. *Nonmonotonic Reasoning: Logical Foundations of Commonsense.* Cambridge University Press, Cambridge, England, 1991.

[Buchanan and Shortliffe, 1984] Bruce Buchanan and Edward H. Shortliffe. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison Wesley, Reading, Massachusetts, 1984.

[Burger and Bhanu, 1992] Wilhelm Burger and Bir Bhanu. *Qualitative Motion Planning.* Kluwer Academic Publishers, Dordrecht, 1992.

[Burgess, 1984] John P. Burgess. Basic tense logic. In Dov Gabbay and Franz Guenther, editors, *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, pages 89–133. D. Reidel Publishing Co., Dordrecht, 1984.

[Burkhard *et al.*, 1998] Hans-Dieter Burkhard, Markus Hannebauer, and Jan Wendler. Belief-desire-intention deliberation in artificial soccer. *The AI Magazine*, 1998(3):87–93, 1998.

[Buszkowski, 1996] Wojciech Buszkowski. Mathematical linguistics and proof theory. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 683–736. Elsevier, Amsterdam, 1996.

[Carlson and Pelletier, 1995] Greg N. Carlson and Francis Jeffrey Pelletier, editors. *The Generic Book.* Chicago University Press, Chicago, IL, 1995.

[Carnap, 1955] Rudolph Carnap. Meaning and synonymy in natural languages. *Philosophical Studies*, 7:33–47, 1955. Reprinted in [Carnap, 1956], pp. 233–247.

[Carnap, 1956] Rudolph Carnap. *Meaning and Necessity.* Chicago University Press, Chicago, 2 edition, 1956. (First edition published in 1947.).

[Casati and Varzi, 1996] Roberto Casati and Achille C. Varzi. *Holes and Other Superficialities.* The MIT Press, Cambridge, Massachusetts, 1996.

[Casati and Varzi, 1999] Roberto Casati and Achille C. Varzi. *Parts and Places: The Structures of Spatial Representation.* The MIT Press, Cambridge, Massachusetts, 1999.

[Chellas, 1975] Brian Chellas. Basic conditional logic. *Journal of Philosophical Logic*, 4(2):133–154, 1975.

[Chen, 1990] Su-Shing Chen, editor. *Advances in Spatial Reasoning, Volume 1.* Ablex, Norwood, New Jersey, 1990.

[Clancey, 1983] William J. Clancey. The epistemology of a rule-based expert system: a framework for explanation. *Artificial Intelligence*, 20:215–251, 1983.

[Clark, 1978] Keith L. Clark. Negation as failure. In H. Gallaire and Jack Minker, editors, *Logic and Data Bases*, pages 293–322. Plenum Press, New York, 1978.

[Clarke, 1981] Bowman L. Clarke. A calculus of individuals based on 'connection'. *Notre Dame Journal of Formal Logic*, 22(3):204–218, 1981.

[Clarke, 1985] Bowman L. Clarke. Individuals and points. *Notre Dame Journal of Formal Logic*, 26(1):61–75, 1985.

[Cohen and Levesque, 1990] Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(3):213–261, 1990.

[Cohn *et al.*, 1997] Anthony G. Cohn, Brandon Bennett, John Gooday, and Nicholas M. Gotts. Representing and reasoning with qualitative spatial relations. In Oliviero Stock, editor, *Spatial and Temporal Reasoning*, pages 97–134. Kluwer Academic Publishers, Dordrecht, 1997.

[Cohn *et al.*, 1998] Anthony G. Cohn, Lenhart Schubert, and Stuart C. Shapiro, editors. *KR'89: Principles of Knowledge Representation and Reasoning.* Morgan Kaufmann, San Francisco, 1998.

[Cohn *et al.*, 2000] Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, editors. *KR2000: Principles of Knowledge Representation and Reasoning.* Morgan Kaufmann, San Francisco, 2000.

[Cohn, 1996] Anthony G. Cohn. Qualitative spatial representation and reasoning techniques. In Gerhard Brewka, Christopher Habel, and Bernhard Nebel, editors, *KI-97, Advances in Artificial Intelligence*, pages 1–30. Springer-Verlag, Berin, 1996.

[Copeland, 1996] B. Jack Copeland. Arthur Prior's life and legacy. In Jack Copeland, editor, *Logic and Reality: Essays on the Legacy of Arthur Prior*, pages 1–40. Oxford University Press, Oxford, 1996.

[Davis, 1988] Martin Davis. Mathematical logic and the origin of modern computers. In Rolf Herkin, editor, *The Universal Turing Machine: A Half-Century Survey*, pages 149–174. Oxford University Press, Oxford, 1988.

[Davis, 1991] Ernest Davis. *Common Sense Reasoning.* Morgan Kaufmann, San Francisco, 1991.

[DeJong and Bennett, 1989] Gerald D. DeJong and Scott W. Bennett. Permissive planning: Extending classical planning to uncertain task domains. *Artificial Intelligence*, 89(1–2):173–217, 1989.

[Delgrande, 1998] James P. Delgrande. Conditional logics for defeasible logics. In Dov M. Gabbay and Philippe Smets, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 2*, pages 135–174. Kluwer Academic Publishers, Dordrecht, 1998.

[Dennett, 1987] Daniel Dennett. Cognitive wheels: The frame problem of AI. In Zenon Pylyshyn, editor, *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, pages 41–64. Ablex Publishing Co., Norwood, New Jersey, 1987.

[Dowty, 1977] David R. Dowty. Toward a semantic analysis of verb aspect and the English 'imperfective' progressive. *Linguistics and Philosophy*, 1(1):45–77, 1977.

[Doyle and Thomason, 1999] Jon Doyle and Richmond H. Thomason. Background to qualitative decision theory. *AI Magazine*, 20(2):55–68, 1999.

[Doyle *et al.*, 1994] Jon Doyle, Erik Sandewall, and Pietro Torasso, editors. *KR'94: Principles of Knowledge Representation and Reasoning.* Morgan Kaufmann, San Francisco, 1994.

[Doyle, 1979] Jon Doyle. A truth maintenance system. *Artificial Intelligence*, 12(3):231–272, 1979.

[Elkan, 1991] Charles Elkan. Reasoning about action in first-order logic. In *Proceedings of the Conference of the Canadian Society for Computational Studies of Intelligence (CSCSI)*, pages 221–227, San Francisco, 1991. Canadian Society for Computational Studies of Intelligence, Morgan Kaufman.

[Elkan, 1995] Charles Elkan. On solving the qualification problem. In *Working Notes of the AAAI Spring Symposium on Extending Theories of Action: Formal Theories and Applications*, Menlo Park, California, 1995. American Association for Artificial Intelligence.

[Fagin *et al.*, 1984] Ronald Fagin, Joseph Y. Halpern, and Moshe Y. Vardi. A model theoretic analysis of knowledge. In *Proceedings of the Twenty-Fifth Annual Symposium on the Foundations of Computer Science*, pages 268–278, 1984.

[Fagin *et al.*, 1995] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge.* The MIT Press, Cambridge, Massachusetts, 1995.

[Fensel *et al.*, 2002] Dieter Fensel, Fausto Giunchiglia, Deborah McGuinness, and Mary-Anne Williams, editors. *KR2002: Principles of Knowledge Representation and Reasoning.* Morgan Kaufmann, San Francisco, California, 2002.

[Fikes, 1996] Richard Fikes. Ontologies: What are they, and where's the research? In Luigia Carlucci Aiello, Jon Doyle, and Stuart Shapiro, editors, *KR'96: Principles of Knowledge Representation and Reasoning*, pages 652–654. Morgan Kaufmann, San Francisco, California, 1996.

[Finger, 1987] Jeffrey J. Finger. *Exploiting Constraints in Design Synthesis.* Ph.D. dissertation, Department of Computer Science, Stanford University, Stanford, California, 1987.

[Fodor, 1987] Jerry A. Fodor. Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In Zenon Pylyshyn, editor, *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, pages 139–149. Ablex Publishing Co., Norwood, New Jersey, 1987.

[Forbus *et al.*, 1991] Kenneth D. Forbus, Paul Nielsen, and Boi Faltings. Qualitative spatial reasoning: The Clock project. *Artificial Intelligence*, 51(1–3):417–471, 1991.

[Forbus, 1988] Kenneth D. Forbus. Qualitative physics: Past, present, and future. In Howard E. Shrobe, editor, *Exploring Artificial Intelligence*, pages 239–296. Morgan Kaufmann, San Mateo, California, 1988.

[Ford and Pylyshyn, 1996] Kenneth M. Ford and Zenon Pylyshyn, editors. *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence.* Ablex Publishing Co., Norwood, New Jersey, 1996.

[Gabbay *et al.*, 1994] Dov Gabbay, Christopher J. Hogger, and J. A. Robinson, editors. *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning.* Oxford University Press, Oxford, 1994.

[Gabbay, 1995] Dov M. Gabbay. Conditional implications and non-monotonic consequence. In Gabriella Crocco, Luis Fari nas del Cerro, and A. Herzig, editors, *Conditionals: From Philosophy to Computer Science*, pages 337–359. Oxford University Press, Oxford, 1995.

[Galton, 1997] Anthony Galton. Space, time, and movement. In Oliviero Stock, editor, *Spatial and Temporal Reasoning*, pages 321–352. Kluwer Academic Publishers, Dordrecht, 1997.

[Gärdenfors and Makinson, 1994] Peter Gärdenfors and David Makinson. Nonmonotonic inferences based on expectations. *Artificial Intelligence*, 65(2):197–245, 1994. Uses selection functions and "expectation relations" to model expectations. Relations between theory revision and nonmonotonic logic.

[Geanakopolos, 1994] John Geanakopolos. Common knowledge. In Robert Aumann and S. Hart, editors, *Handbook of Game Theory, Volume 2*, chapter 40. Elsevier Science Publishers, Amsterdam, 1994.

[Geffner, 1990] Hector Geffner. Causal theories of nonmonotonic reasoning. In Thomas Dietterich and William Swartout, editors, *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 524–530, Menlo Park, CA, 1990. American Association for Artificial Intelligence, AAAI Press.

[Geffner, 1992] Hector Geffner. *Default Reasoning: Causal and Conditional Theories.* MIT Press, Cambridge, Massachusetts, 1992.

[Gelfond and Lifschitz, 1998] Michael Gelfond and Vladimir Lifschitz. Action Languages. *Electronic Transactions on AI*, 3, 1998. Available at http://www.ep.liu.se/rs/cis/1998/016/.

[Genesereth and Nilsson, 1987] Michael Genesereth and Nils J. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann, San Mateo, California, 1987.

[Ginsberg, 1987] Matthew L. Ginsberg, editor. *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann, Los Altos, California, 1987. (Out of print.).

[Giunchiglia and Lifschitz, 1998] Enrico Giunchiglia and Vladimir Lifschitz. An action language based on causal explanation. In Ted Senator and Bruce Buchanan, editors, *Proceedings of the Fourteenth National Conference on Artificial Intelligence and the Ninth Innovative Applications of Artificial Intelligence Conference*, pages 623–628, Menlo Park, California, 1998. American Association for Artificial Intelligence, AAAI Press.

[Giunchiglia *et al.*, 1997] Enrico Giunchiglia, G. Neelakantan Kartha, and Vladimir Lifschitz. Representing action: Indeterminacy and ramifications. *Artificial Intelligence*, 95(2):409–438, 1997.

[Glasgow *et al.*, 1995] Janice Glasgow, N. Hari Narayanan, and B. Chandrasekaran, editors. *Diagrammatic Reasoning*. The MIT Press, Cambridge, Massachusetts, 1995.

[Golden and Weld, 1996] Keith Golden and Daniel Weld. Representing sensing actions: The middle ground revisited. In Luigia Carlucci Aiello, Jon Doyle, and Stuart Shapiro, editors, *KR'96: Principles of Knowledge Representation and Reasoning*, pages 174–185. Morgan Kaufmann, San Francisco, California, 1996.

[Goodman, 1946] Nelson Goodman. *Fact, Fiction and Forecast*. Harvard University Press, fourth edition, 1946.

[Gotts, 1994] N.M. Gotts. How far can we 'C'? defining a doughnut using connection alone. In Jon Doyle, Erik Sandewall, and Pietro Torasso, editors, *KR'94: Principles of Knowledge Representation and Reasoning*, pages 246–257. Morgan Kaufmann, San Francisco, California, 1994.

[Gotts, 1996] N.M. Gotts. Topology from a single primitive relation: Defining topological properties and relations in terms of connection. Technical Report 96.24, School of Computer Studies, University of Leeds, Leeds, 1996.

[Guha, 1991] Ramanathan V. Guha. Contexts: a formalization and some applications. Technical Report STAN-CS-91-1399, Stanford Computer Science Department, Stanford, California, 1991.

[Gustaffson and Doherty, 1996] Joakim Gustaffson and Patrick Doherty. Embracing occlusion in specifying the indirect effects of actions. In Luigia Carlucci Aiello, Jon Doyle, and Stuart Shapiro, editors, *KR'96: Principles of Knowledge Representation and Reasoning*, pages 87–98. Morgan Kaufmann, San Francisco, California, 1996.

[Halpern and Moses, 1985] Joseph Y. Halpern and Yoram Moses. Towards a theory of knowledge and ignorance. In Krzysztof R. Apt, editor, *Logics and Models of Concurrent Systems*, pages 459–476. Springer-Verlag, Berlin, 1985.

[Halpern, 1986] Joseph Y. Halpern, editor. *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the First Conference (TARK 1986).* Morgan Kaufmann Publishers, Inc., Los Altos, California, 1986.

[Hammer, 1995] Eric M. Hammer. *Logic and Visual Information.* CSLI Publications, Stanford, California, 1995.

[Hanks and McDermott, 1985] Steven Hanks and Drew McDermott. Temporal reasoning and default logics. Technical Report YALEU/CSD/RR#430, Department of Computer Science, Yale University, New Haven, Connecticut, 1985.

[Hanks and McDermott, 1986] Steven Hanks and Drew McDermott. Default reasoning, nonmonotonic logics and the frame problem. In Tom Kehler and Stan Rosenschein, editors, *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 328–333, Los Altos, California, 1986. American Association for Artificial Intelligence, Morgan Kaufmann.

[Hanks and McDermott, 1987] Steven Hanks and Drew McDermott. Non-monotonic logics and temporal projection. *Artificial Intelligence*, 33(3):379–412, 1987.

[Haugeland, 1981] John Haugeland. Semantic engines: An introduction to mind design. In John Haugeland, editor, *Mind Design*, pages 1–34. The MIT Press, Cambridge, Massachusetts, 1981.

[Haugh, 1987] Brian Haugh. Simple causal minimization for temporal persistence and projection. In Kenneth Forbus and Howard Shrobe, editors, *Proceedings of the Seventh National Conference on Artificial Intelligence*, pages 218–223, Menlo Park, California, 1987. American Association for Artificial Intelligence, AAAI Press.

[Hintikka, 1962] Jaakko Hintikka. *Knowledge and Belief.* Cornell University Press, Ithaca, New York, 1962.

[Hintikka, 1970] Jaakko Hintikka. Objects of knowledge and belief: Aquaintances and public figures. *The Journal of Philosophy*, 67(21):869–883, 1970.

[Hintikka, 1986] Jaakko Hintikka. Reasoning about knowledge in philosophy: The paradigm of epistemic logic. In Joseph Y. Halpern, editor, *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the First Conference*, pages 63–80, Los Altos, California, 1986. Morgan Kaufmann Publishers, Inc.

[Hobbs *et al.*, 1993] Jerry Hobbs, Mark Stickel, Douglas Appelt, and Paul Martin. Interpretation as abduction. *Artificial Intelligence*, 63(1–2):69–142, 1993.

[Horty, 1994] John F. Horty. Some direct theories of nonmonotonic inheritance. In Dov Gabbay, Christopher J. Hogger, and J. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*, pages 111–187. Oxford University Press, 1994.

[Horty, 2001] John F. Horty. *Agency and Deontic Logic*. Oxford University Press, Oxford, 2001.

[Israel, 1991] David J. Israel. A short sketch of the life and career of john McCarthy. In Vladimir Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation: Papers in Honor of John McCarthy*. Academic Press, San Diego, California, 1991.

[Iwasaki and Simon, 1986] Yumi Iwasaki and Herbert Simon. Causality in device behavior. *Artificial Intelligence*, 29(1):3–32, 1986.

[Kapur and Mundy, 1988] Deepak Kapur and Joseph L. Mundy. Geometric reasoning and artificial intelligence: Introduction to the special volume. *Artificial Intelligence*, 37(1–3):1–11, 1988.

[Konolige and Pollack, 1993] Kurt Konolige and Martha Pollack. A representationalist theory of intention. In Ruzena Bajcsy, editor, *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, San Mateo, California, 1993. Morgan Kaufmann.

[Konolige, 1986] Kurt Konolige. What awareness isn't: A sentential view of implicit and explicit belief. In Joseph Y. Halpern, editor, *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the First Conference*, pages 241–250, Los Altos, California, 1986. Morgan Kaufmann Publishers, Inc.

[Konolige, 1988] Kurt Konolige. On the relation between default and autoepistemic logic. *Artificial Intelligence*, 35(3):343–382, 1988. (See also errata, Artificial Intelligence **41**(1): 115.).

[Konolige, 1994] Kurt Konolige. Autoepistemic logic. In Dov Gabbay, Christopher J. Hogger, and J. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*, pages 217–295. Oxford University Press, Oxford, 1994.

[Kosslyn, 1990] Stephen Michael Kosslyn. Visual cognition: Introduction. In Daniel N. Osherson and Howard Lasnik, editors, *An Invitation to Cognitive Science. Volume 2: Visual Cognition and Action*, pages 3–4. The MIT Press, Cambridge, Massachusetts, 1990.

[Kowalski and Sergot, 1986] Robert A. Kowalski and Marek J. Sergot. A logic-based calculus of events. *New Generation Computing*, 4:67–95, 1986.

[Krifka *et al.*, 1995] Manfred Krifka, Francis Jeffrey Pelletier, Gregory Carlson, Alice ter Meulen, Gennaro Chierchia, and Godehard Link. Genericity: An introduction. In Gregory Carlson and Francis Jeffrey Pelletier, editors, *The Generic Book*, pages 1–124. Chicago University Press, Chicago, IL, 1995.

[Kuipers, 1993] Benjamin Kuipers. Reasoning with qualitative models. *Artificial Intelligence*, 59(1–2):125–132, 1993.

[Lakemeyer, 1997] Gerhard Lakemeyer. Limited reasoning in first-order knowledge bases. *Artificial Intelligence*, 71(2):213–255, 1997.

[Lascarides, 1992] Alex Lascarides. The progressive and the imperfective paradox. *Synthèse*, 87(6):401–447, 1992.

[Laux and Wansing, 1995] Armin Laux and Heinrich Wansing, editors. *Knowledge and Belief in Philosophy and Artificial Intelligence.* Akedemie Verlag, Berlin, 1995.

[Lehmann and Magidor, 1992] Daniel Lehmann and Menachem Magidor. What does a conditional knowledge base entail? *Artificial intelligence*, 55(1):1–60, 1992.

[Lenat and Guha, 1989] Douglas B. Lenat and R.V. Guha. *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project.* Addison-Wesley Publishing Company, Reading, Massachusetts, 1989.

[Leśniewski, 1916] Stanisław Leśniewski. Podstawy ogólnej teorii mnogości I, 1916. English Title: "Foundations of a general set theory I.".

[Levesque and Lakemeyer, 2000] Hector Levesque and Gerhard Lakemeyer. *The Logic of Knowledge Bases.* The MIT Press, Cambridge, Massachusetts, 2000.

[Levesque, 1984] Hector J. Levesque. A logic of implicit and explicit belief. In *Proceedings of the Fourth National Conference on Artificial Intelligence*, pages 198–202. American Association for Artificial Intelligence, 1984.

[Levesque, 1987] Hector J. Levesque. Taking issue: Guest editor's introduction. *Computational Intelligence*, 3(3):149–150, 1987.

[Levy, 2000] Alon Y. Levy. Logic-based techniques in data integration. In Jack Minker, editor, *Logic-Based Artificial Intelligence*, pages 575–595. Kluwer Academic Publishers, Dordrecht, 2000.

[Lifschitz, 1987] Vladimir Lifschitz. Formal theories of action: Preliminary report. In John McDermott, editor, *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, Los Altos, California, 1987. Morgan Kaufmann.

[Lifschitz, 1990a] Vladimir Lifschitz, editor. *Formalizing Common Sense: Papers by John McCarthy.* Ablex Publishing Corporation, Norwood, New Jersey, 1990.

[Lifschitz, 1990b] Vladimir Lifschitz. Understanding common sense: McCarthy's research in artificial intelligence. In Vladimir Lifschitz, editor, *Formalizing Common Sense: Papers by John McCarthy*, pages 1–8. Ablex Publishing Corporation, Norwood, New Jersey, 1990.

[Lifschitz, 1997] Vladimir Lifschitz. On the logic of causal explanation. *Artificial Intelligence*, 96(2):451–465, 1997.

[Lifschitz, 1998] Vladimir Lifschitz. Situation calculus and causal logic. In Anthony G. Cohn, Lenhart Schubert, and Stuart C. Shapiro, editors, *KR'98: Principles of Knowledge Representation and Reasoning*, pages 536–546. Morgan Kaufmann, San Francisco, California, 1998.

[Lin, 1995] Fangzhen Lin. Embracing causality in specifying the indirect effects of actions. In Chris Mellish, editor, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1985–1991, San Francisco, 1995. Morgan Kaufmann.

[Lormand, 1996] Eric Lormand. The holorobophobe's dilemma. In Kenneth M. Ford and Zenon Pylyshyn, editors, *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*, pages 61–88. Ablex Publishing Co., Norwood, New Jersey, 1996.

[Łukasiewicz, 1970] Jan Łukasiewicz. *Jan Łukasiewicz, Selected Writings.* North-Holland, 1970. Edited by Ludwik Borowski.

[Lukaszewicz, 1990] Witold Lukaszewicz. *Non-Monotonic Reasoning: Formalization of Commonsense Reasoning.* Ellis Horwood, New York, 1990.

[Marek and Truszczyński, 1989] Wictor Marek and Mirosław Truszczyński. Relating autoepistemic and default logics. In Ronald J. Brachman, Hector J. Levesque, and Raymond Reiter, editors, *KR'89: Principles of Knowledge Representation and Reasoning*, pages 276–288. Morgan Kaufmann, San Mateo, California, 1989.

[Marek and Truszczyński, 1991] Victor Marek and Mirosław Truszczyński. Autoepistemic logic. *Journal of the Association for Computing Machinery*, 38(3):588–619, 1991.

[Marek and Truszczyński, 1994] Wictor Marek and Mirosław Truszczyński. *Nonmonotonic Logic: Context-Dependent Reasoning.* Springer-Verlag, Berlin, 1994.

[McCain and Turner, 1995] Norman McCain and Hudson Turner. A causal theory of ramifications and qualifications. In Chris Mellish, editor, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1978–1984, San Francisco, 1995. Morgan Kaufmann.

[McCain and Turner, 1997] Norman McCain and Hudson Turner. Causal theories of action and change. In Howard Shrobe and Ted Senator, editors, *Proceedings of the Thirteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference*, pages 460–465, Menlo Park, California, 1997. American Association for Artificial Intelligence, AAAI Press.

[McCarthy and Buvač, 1998] John McCarthy and Saša Buvač. Formalizing context (expanded notes). In Atocha Aliseda, Rob van Glabbeek, and Dag Westerståhl, editors, *Computing Natural Language*, pages 13–50. CSLI Publications, Stanford, California, 1998.

[McCarthy and Hayes, 1969] John McCarthy and Patrick J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, Edinburgh, 1969.

[McCarthy, 1959] John McCarthy. Programs with common sense. In *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, pages 75–91, London, 1959. Her Majesty's Stationary Office.

[McCarthy, 1979] John McCarthy. First order theories of individual concepts and propositions. In J.E. Hayes, D. Mitchie, and L.I. Mikulich, editors, *Machine Intelligence 9*, pages 129–148. Ellis Horwood, Chichester, England, 1979.

[McCarthy, 1980] John McCarthy. Circumscription: A form of non-monotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980.

[McCarthy, 1983] John McCarthy. Situations, actions, and causal laws. Technical Report Memo 2, Stanford Artificial Intelligence Project, Stanford University, 1983.

[McCarthy, 1986] John McCarthy. Applications of circumscription to formalizing common sense knowledge. *Artificial Intelligence*, 13:27–39, 1986.

[McCarthy, 1987] John McCarthy. Epistemological problems of artificial intelligence. In Matthew L. Ginsberg, editor, *Readings in Nonmonotonic Reasoning*, pages 46–55. Morgan Kaufmann, Los Altos, California, 1987.

[McCarthy, 1993a] John McCarthy. History of circumscription. *Artificial Intelligence*, 59:23–26, 1993.

[McCarthy, 1993b] John McCarthy. Notes on formalizing contexts. In Ruzena Bajcsy, editor, *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 555–560, San Mateo, California, 1993. Morgan Kaufmann.

[McCarthy, 1999] John McCarthy. Elaboration tolerance, 1999. http://www-formal.stanford.edu/jmc/elaboration.html.

[McDermott and Doyle, 1980] Drew McDermott and Jon Doyle. Non-monotonic logic I. *Artificial Intelligence*, 13:41–72, 1980.

[McDermott, 1982] Drew McDermott. Nonmonotonic logic II: Nonmonotonic modal theories. *Journal of the Association for Computing Machinery*, 29(1):33–57, 1982.

[McDermott, 1987] Drew McDermott. Critique of pure reason. *Computational Intelligence*, 3(3):149–160, 1987.

[Meyer and van der Hoek, 1995] John-Jules Ch. Meyer and Wiebe van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge University Press, Cambridge, 1995.

[Minker, 1997] Jack Minker. Logic and databases: Past, present and future. *AI Magazine*, 18(3):21–47, 1997.

[Minker, 2000a] Jack Minker. Introduction to logic-based artificial intelligence. In Jack Minker, editor, *Logic-Based Artificial Intelligence*, pages 3–33. Kluwer Academic Publishers, Dordrecht, 2000.

[Minker, 2000b] Jack Minker, editor. *Logic-Based Artificial Intelligence.* Kluwer Academic Publishers, Dordrecht, 2000.

[Minsky, 1974] Marvin Minsky. A framework for representing knowledge. Technical Report 306, Artificial Intelligence Laboratory, MIT, 1974. Republished in several places, including [Haugeland, 1981].

[Moore, 1985] Robert C. Moore. A formal theory of knowledge and action. In Jerry R. Hobbs and Robert C. Moore, editors, *Formal Theories of the Commonsense World*, pages 319–358. Ablex Publishing Corporation, Norwood, New Jersey, 1985.

[Moore, 1993] Robert C. Moore. Autoepistemic logic revisited. *Artificial Intelligence*, 59(1–2):27–30, 1993.

[Moore, 1995a] Johanna Moore. *Participating in Explanatory Dialogues.* The MIT Press, 1995.

[Moore, 1995b] Robert C. Moore. *Logic and Representation.* Cambridge University Press, Cambridge, England, 1995.

[Morgenstern and Stein, 1994] Leora Morgenstern and Lynn Stein. Motivated action theory: a formal theory of causal reasoning. *Artificial Intelligence*, 71(1):1–42, 1994.

[Morgenstern, 1996] Leora Morgenstern. The problem with solutions to the frame problem. In Kenneth M. Ford and Zenon Pylyshyn, editors, *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*, pages 99–133. Ablex Publishing Co., Norwood, New Jersey, 1996.

[Nakashima *et al.*, 1997] Hideyuki Nakashima, Hitoshi Matsubara, and Ichiro Osawa. Causality as a key to the frame problem. *Artificial Intelligence*, 91(1):37–50, 1997.

[Nebel *et al.*, 1992] Bernhard Nebel, Charles Rich, and William Swartout, editors. *KR': Principles of Knowledge Representation and Reasoning.* Morgan Kaufmann, San Francisco, 1992.

[Nilsson, 1991] Nils J. Nilsson. Logic and artificial intelligence. *Artificial Intelligence*, 47(1–3):31–56, 1991.

[Nilsson, 1995] Nils J. Nilsson. Eye on the prize, 1995. Available at http://robotics.Stanford.EDU/~nilsson/.

[Ohrstrom and Hasle, 1995] Peter Ohrstrom and Per F.V. Hasle. *Temporal Logic from Ancient Ideas to Artificial Intelligence.* Kluwer Academic Publishers, Dordrecht, 1995.

[Osborne and Rubenstein, 1994] Martin J. Osborne and Ariel Rubenstein. *A Course in Game Theory.* The MIT Press, Cambridge, Massachusetts, 1994.

[Osherson and Lasnik, 1990] Daniel N. Osherson and Howard Lasnik, editors. *An Invitation to Cognitive Science. Volume 2: Visual Cognition and Action.* The MIT Press, Cambridge, Massachusetts, 1990.

[Pearl, 1994] Judea Pearl. From Adams' conditionals to default expressions, causal conditionals, and counterfactuals. In Ellery Eells and Brian Skyrms, editors, *Probability and Conditionals: Belief Revision and Rational Decision*, pages 47–74. Cambridge University Press, Cambridge, England, 1994.

[Pearl, 2000] Judea Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, Cambridge, England, 2000.

[Perlis, 1985] Donald Perlis. Languages with self-reference I: Foundations. *Artificial Intelligence*, 25:301–322, 1985.

[Pirri and Finzi, 1999] Fiora Pirri and Alberto Finzi. A preliminary approach to perception in theory of agents. In Michael Thielscher, editor, *Proceedings of the IJCAI-99 Workshop on Nonmonotonic Reasoning, Action and Change*, pages 49–56, Murray Hill, New Jersey, 1999. IJCAI, International Joint Conference on Artificial Intelligence.

[Pollack, 1992] Martha Pollack. The uses of plans. *Artificial Intelligence*, 57(1):43–68, 1992.

[Pollock, 1995] John L. Pollock. *Cognitive Carpentry: A Manual for How to Build a Person.* The MIT Press, Cambridge, Massachusetts, 1995.

[Prior, 1956] Arthur Prior. *Time and Modality.* Oxford University Press, Oxford, 1956.

[Prior, 1967] Arthur Prior. *Past, Present and Future.* Oxford University Press, Oxford, 1967.

[Prior, 1968] Arthur Prior. *Papers on Time and Tense.* Oxford University Press, Oxford, 1968.

[Pylyshyn, 1987] Zenon Pylyshyn, editor. *The Robot's Dilemma: The Frame Problem in Artificial Intelligence.* Ablex Publishing Co., Norwood, New Jersey, 1987.

[Quine, 1960] Willard V.O Quine. Variables explained away. In Willard V. Quine, editor, *Selected Logic Papers*, pages 227–235. Harvard University Press, Cambridge, Massachusetts, 1960.

[Rao and Georgeff, 1991] Anand S. Rao and Michael P. Georgeff. Modeling rational agents within a BDI-architecture. In James F. Allen, Richard Fikes, and Erik Sandewall, editors, *KR'91: Principles of Knowledge Representation and Reasoning*, pages 473–484. Morgan Kaufmann, San Mateo, California, 1991.

[Reiter, 1978] Raymond Reiter. On closed world data bases. In H. Gallaire and J. Minker, editors, *Logic and Data Bases*, pages 55–76. Plenum Press, New York, 1978.

[Reiter, 1980] Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–32, 1980.

[Reiter, 1993] Raymond Reiter. Proving properties of states in the situation calculus. *Artificial Intelligence*, 64:337–351, 1993.

[Reiter, 2001] Raymond Reiter. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems.* The MIT Press, Cambridge, Massachusetts, 2001.

[Renz and Nebel, 1999] Jochen Renz and Bernhard Nebel. On the complexity of qualitative spatial reasoning: A maximal tractable fragment of the region connection calculus. *Artificial Intelligence*, 108(1–2):69–123, 1999.

[Rosenschein and Kaelbling, 1995] Stanley J. Rosenschein and Leslie Pack Kaelbling. A situated view of representation and control. *Artificial Intelligence*, 73(1–2):149–173, 1995.

[Rosenschein, 1989] Stanley J. Rosenschein. Synthesizing information-tracking automata from environment descriptions. In Ronald J. Brachman, Hector J. Levesque, and Raymond Reiter, editors, *KR'89: Principles of Knowledge Representation and Reasoning*, pages 386–393. Morgan Kaufmann, San Mateo, California, 1989.

[Rounds, 1996] William C. Rounds. Feature logics. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 475–533. Elsevier, Amsterdam, 1996.

[Russell and Wefald, 1991] Stuart J. Russell and Eric Wefald. *Do the Right Thing.* The MIT Press, Cambridge, Massachusetts, 1991.

[Russell, 1957] Bertrand Russell. *Mysticism and Logic.* Doubleday, Garden City, New York, 1957.

[Sadek, 1992] M.D. Sadek. A study in the logic of intention. In Bernhard Nebel, Charles Rich, and William Swartout, editors, *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, pages 462–473. Morgan Kaufmann, San Mateo, California, 1992.

[Sandewall, 1972] Eric Sandewall. An approach to the frame problem, and its implementation. In D. Michie and B. Meltzer, editors, *Machine Intelligence 7*, pages 195–204. Edinburgh University Press, 1972.

[Sandewall, 1994] Erik Sandewall. *Features and Fluents: A Systematic Approach to the Representation of Knowledge About Dynamical Systems.* Oxford University Press, Oxford, 1994.

[Schlechta, 1997] Karl Schlechta. *Nonmonotonic Logics.* Springer-Verlag, Berin, 1997.

[Schubert, 1990] Lenhart Schubert. Monotonic solution of the frame problem in the situation calculus; an efficient method for worlds with fully specified actions. In Henry Kyburg, Ronald Loui, and Greg Carlson, editors, *Knowledge Representation and Defeasible Reasoning*, pages 23–67. Kluwer Academic Publishers, Dordrecht, 1990.

[Seligman and Moss, 1996] Jerry Seligman and Lawrence S. Moss. Situation theory. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 239–307. Elsevier, Amsterdam, 1996.

[Shanahan, 1997] Murray Shanahan. *Solving the Frame Problem.* The MIT Press, Cambridge, Massachusetts, 1997.

[Shieber, 1992] Stuart Shieber. *Constraint-Based Grammar Formalisms: Parsing and Type Inference for Natural and Computer Languages.* The MIT Press, Cambridge, Massachusetts, 1992.

[Shoham, 1988] Yoav Shoham. *Reasoning About Change: Time and Causation From the Standpoint of Artificial Intelligence.* The MIT Press, Cambridge, Massachusetts, 1988.

[Simon, 1952] Herbert Simon. On the definition of the causal relation. *The Journal of Philosophy*, 49:517–528, 1952.

[Simon, 1966] Herbert Simon. On reasoning about action. Technical Report Complex Information Processing Paper #87, Carnegie Institute of Technology, Pittsburgh, Pennsylvania, 1966.

[Simon, 1977] Herbert A. Simon. *Models of Discovery.* D. Reidel Publishing Co., Dordrecht, 1977.

[Simon, 1982a] Herbert A. Simon. *Models of Bounded Rationality, Volume 1.* The MIT Press, Cambridge, Massachusetts, 1982.

[Simon, 1982b] Herbert A. Simon. *Models of Bounded Rationality, Volume 2.* The MIT Press, Cambridge, Massachusetts, 1982.

[Simons, 1987] Peter Simons. *Parts: A Study in Ontology.* Oxford University Press, Oxford, 1987.

[Stalnaker, 1993] Robert C. Stalnaker. A note on non-monotonic modal logic. *Artificial Intelligence*, 64(2):183–196, 1993. Widely circulated in manuscipt form, 1980 to 1992.

[Steedman, 1995] Mark Steedman. Dynamic semantics for tense and aspect. In Chris Mellish, editor, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1292–1298, San Francisco, 1995. Morgan Kaufmann.

[Steedman, 1998] Mark Steedman. The productions of time. Unpublished manuscript, University of Edinburgh. Available from http://www.cogsci.ed.ac.uk/ steedman/papers.html., 1998.

[Stefik, 1995] Mark J. Stefik. *An Introduction to Knowledge Systems.* Morgan Kaufmann, San Francisco, 1995.

[Stock, 1997] Oliviero Stock, editor. *Spatial and Temporal Reasoning.* Kluwer Academic Publishers, Dordrecht, 1997.

[Stone, 1998] Matthew Stone. *Modality in Dialogue: Planning, Pragmatics and Computation.* Ph.D. dissertation, Computer Science Department, University of Pennsylvania, Philadelphia, Pennsylvania, 1998.

[Thielscher, 1989] Michael Thielscher. Ramification and causality. *Artificial Intelligence*, 89(1–2):317–364, 1989.

[Thielscher, 1996] Michael Thielscher. Causality and the qualification problem. In Luigia Carlucci Aiello, Jon Doyle, and Stuart Shapiro, editors, *KR'96: Principles of Knowledge Representation and Reasoning*, pages 51–62. Morgan Kaufmann, San Francisco, California, 1996.

[Thielscher, 2000] Michael Thielscher. Representing the knowledge of a robot. In Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, editors, *KR2000: Principles of Knowledge Representation and Reasoning*, pages 109–120, San Francisco, 2000. Morgan Kaufmann.

[Thomason, 1992] Richmond H. Thomason. NETL and subsequent path-based inheritance theories. In Fritz Lehmann, editor, *Semantic Networks in Artificial Intelligence*, pages 179–204. Pergamon Press, Oxford, 1992.

[Turner, 1999] Hudson Turner. A logic of universal causation. *Artificial Intelligence*, 113(1–2):87–123, 1999.

[van Benthem and ter Meulen, 1996] Johan van Benthem and Alice ter Meulen. *Handbook of Logic and Language*. Elsevier Science Publishers, Amsterdam, 1996.

[van Benthem, 1983] Johan van Benthem. *The Logic of Time*. D. Reidel Publishing Company, Dordrecht, 1983.

[van der Torre, 1997] Leendert W.N. van der Torre. *Reasoning about Obligations: Defeasibility in Preference-Based Deontic Logic*. Thesis Publishers, Amsterdam, 1997.

[von Wright, 1983] Georg Henrik von Wright. *Practical Reason: Philosophical Papers, Volume 1*. Cornell University Press, Ithaca, 1983.

[Weld and de Kleer, 1990] Daniel S. Weld and Johan de Kleer, editors. *Qualitative Reasoning about Physical Systems*. Morgan Kaufmann, San Mateo, California, 1990.

[Wilson, 1998] Randall H. Wilson. Geometric reasoning about assembly tools. *Artificial Intelligence*, 98(1–2):237–279, 1998.

[Wobcke *et al.*, 1998] Wayne Wobcke, Maurice Pagnucco, and C. Zhang, editors. *Agents and Multi-Agent Systems—Formalisms, Methodologies, and Applications*. Springer-Verlag, Berlin, 1998.

[Wolter and Zakharyaschev, 2000] Frank Wolter and Michael Zakharyaschev. Spatio-temporal representation and reasoning based on RCC-8. In Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, editors, *KR2000: Principles of Knowledge Representation and Reasoning*, pages 3–14, San Francisco, 2000. Morgan Kaufmann.

[Woods and Schmolze, 1992] William A. Woods and James G. Schmolze. The KL-ONE family. In Fritz Lehmann, editor, *Semantic Networks in Artificial Intelligence*, pages 133–177. Pergamon Press, Oxford, 1992.

[Yeap and Jeffries, 1999] Wai K. Yeap and Margaret E. Jeffries. Computing a representation of the local environment. *Artificial Intelligence*, 107(2):265–301, 1999.