Chapter 1

Introduction

One of the primary goals of artificial intelligence (AI) has always been to make machines that are capable of making good decisions. For example, an ideal AI medical diagnosis and treatment program would need to decide what possible diseases to consider as explanations for the patient's symptoms, what tests to administer to gain more information about the patient's condition, and what medical procedures to perform to alleviate the patient's condition. Or, on a somewhat less critical level, a good chess-playing program would need to decide which lines of play to consider, how much time to spend studying the different possible alternatives, and finally what move to make.

But good decisions are not easy to come by. Economists have studied human decision-making for many years in an effort to understand the economic decision-making behavior of individuals and groups. Drawing from the philosophy of utilitarianism and the mathematics of probability, they have devised an idealized theory of optimal decision-making, called decision theory (DT). But the advent of decision theory has not suddenly made human decision-making an easy thing, for several reasons. One problem is that, although decision theory can help you make decisions when your problem involves unavoidable uncertainties, such as the outcome of a dice roll, it does not directly tell you how to cope with avoidable uncertainties, where you could gain more information before making the decision. You can try to apply decision theory recursively to the problem of deciding what information to gather and how, but this leads to the second problem of decision theory, which is that thoroughly analyzing a decision problem in accordance with the theory can be an arbitrarily long and complex and time-consuming process, so that it would make more sense to just do a simpler, but less accurate analysis.

This is the conundrum of limited rationality. Our brains, and our computers, do not have unlimited speed when reasoning or computing. And neither do we have unlimited time to make our decisions. But using decision theory to its fullest extent can require an

indefinitely-large amount of computation. So, paradoxically, it can be a bad decision to try to make a very good decision. We can never know that applying the theory in any particular case is a better decision than not applying it and just guessing instead. Decision theory really has nothing to say about how to resolve this issue. It does not answer the question of its own use by limited entities such as ourselves and our creations, the computers.

Perhaps partly because of this difficulty, and also because of the extreme mathematical complexity of the more advanced decision-theoretic methods, researchers in AI have not typically used decision theory extensively in the design or operation of their programs. In fact, for many years decision theory (and the foundational disciplines such as probability lying behind it) were almost completely ignored by the AI community. Instead, theoretical AI has focused on a variety of other mathematical frameworks that eschewed the use of quantitative, numeric entities such as probabilities and utilities, and dealt instead with symbolic, discrete, logical entities such as beliefs, plans, and preferences.[1] In the meantime, the more applied, empirical segments of the AI community did not use decision theory either; they pursued other methods in essentially a trial-and-error fashion, taking whatever techniques seemed to work well and refining and developing them, regardless of whether the technique had any kind of deeper theoretical basis or justification. As a result, it is really not well understood why these techniques worked (when they did), and so it was difficult to gain any deeper understanding from writing a program to solve one problem (such as playing chess) that would help in writing a program to solve a different problem (such as clinical decision making).

So, perhaps because of this difficulty of generalization and reuse of the existing techniques, in recent years AI researchers have begun to take another look at foundational theories such as decision theory, despite those theories' problems. The medical AI community has been using decision theory for some time (e.g., [Gorry-et-al-73], [Schwartz-et-al-73]. See [Pauker-Eckman-92] for a brief overview), and now AI researchers are even applying DT to computer game-playing, an area considered by many to be a core testbed for AI techniques, although it is questionable how much transference of ideas there has

1. A somewhat questionable focus, considering that the sciences seem to have always moved from the qualitative to the quantitative, and the quantitative methods, in my opinion, have yielded the lion's share of science's success and power. But perhaps it was just too early in the development of AI for quantitative theories to be fruitful.

been from the specialized realm of game playing programs to other areas of AI. Still, the successful demonstration of these general techniques in game playing could serve to excite many people about those techniques, and lead to their adoption in other domains as a foundation for further progress.

In the work reported in this thesis, I have performed preliminary investigations in several areas related to the use of decision theory in AI. First, I have studied the traditional computer game playing methods and their evolution towards the newest approaches, which make extensive use of decision theory. In this thesis I describe the evolution of the different approaches, and report on the results of new experiments I performed to assess the accuracy of the hypotheses underlying several of them. The recent use of decision theory in game-playing has met with some success, and more importantly, the ability to transfer ideas to the solution of problems other than game playing seems to be much greater with the decision-theoretic methods than with the earlier game playing techniques. However, game-playing algorithms still suffer from a lack of generality, so I have also worked on improving the generalizability of game programs by allowing them to deal with the playing problem on a more abstract level. I report here on my investigations of "abstract search" algorithms. Second, in their attempt to apply decision theory effectively, computer game-playing (CGP) researchers have worked on alleviating decision theory's limited rationality problem. In an effort to understand the generality of their solutions, I have done preliminary investigations of how to transfer the CGP limited-rationality techniques to another decision problem, that of medical decision making. I have also investigated issues concerning the formalization of limited rationality theory, and its possible development into a more practical successor to decision theory. These, too, appear to be promising lines for further study. A primary conclusion of this thesis will be that the right kind of limited rationality theory will involve abstraction in a crucial way.

Properly using the tools of probability and decision theory is not easy, and a future theory of limited rationality would likely be even harder to apply. But from what I have seen so far, these quantitative, foundational theories seem to be a good basis upon which to develop programs that successfully solve decision problems. I believe that eventually, like other sciences, artificial intelligence will make a transition to a state where its theoretical models and empirical methods are precise, stable and quantitative. For AI is really just

another natural science, one that studies a phenomenon—intelligent thought—that occurs naturally, in humans, if nowhere else. Just as physicists learned about universal laws of motion, electricity, etc., through observations of very simplified physical systems in the laboratory,[1] so too, AI can be seen as attempting to discover universal laws of thought by studying simplified forms of reasoning within the "laboratory" of the computer.

1.1 Structure of This Thesis

Chapter 2, "Evolution of Computer Game Playing Techniques," begins with a discussion of the evolution of decision-theoretic methods and the concurrent development of automated reasoning algorithms, particularly for playing games. I report here for the first time some results of experiments performed by myself and a colleague on some of these algorithms. Interestingly, decision theory and game theory historically share a common genesis, but the applied, computational side of game theory then diverged from decision theory, and has only recently begun to bring decision theory back into game-playing with recent attempts to use it at the "meta" level to control the computational reasoning itself.

These recent developments resurrect the old issue of the limited nature of rationality, and the need for a successor of decision theory that could accommodate it. These matters are discussed in Chapter 3, "Limited Rationality," along with some of the preliminary attempts by the author and others to devise such a theory. Many issues remain to be resolved in this area, and they appear very difficult to surmount.

However, regardless of whether a good theory of limited rationality is attained that could form a coherent basis for game-playing programs and other reasoning programs to control their own computational reasoning, the question remains as to whether the whole traditional state-space search oriented approach to game playing and most other kinds of automated reasoning is really appropriate. Many researchers have noted that a more general and efficient method based on concepts such as patterns, "chunking," or abstraction might be possible to develop. Much theory and practical methodology in this area remains to be developed, but Chapter 4, "Abstract State-Space Search," describes some of the preliminary developments in this direction.

---

1. For example, Galileo's experiments on gravity that were conducted by rolling objects down inclined planes.

So, where is this all going? Chapter 5, "Summary and Conclusions," sums up and presents my vision of some of the AI research of the future that will be most fruitful for myself and others to explore.

1.2 Related W ork

A more detailed overview of classical computer game-playing than the one presented in Chapter 2 can be found in [Newborn-89]. The book [Marsland-Schaeffer-90] contains a good selection of fairly recent readings, and the proceedings of the Spring 1993 AAAI symposium on "Games: Planning and Learning," [Games-93], contains a wide spectrum of some of the most recent research.

A good survey of some decision-theoretic methods for control of computation is provided in [Dean-91]. The papers [Russell-Wefald-91] and [Russell-et-al-93] both contain very useful sections on the history of limited rationality in AI and elsewhere, with a number of good pointers to the literature. The working notes of the 1989 symposium on AI and Limited Rationality, [AILR-89], is a good selection of fairly recent work on limited rationality within the AI community, although it is difficult to obtain.

The body of literature in AI that is relevant to the topic of Chapter 4, "Abstract State-Space Search," is large, and not adequately explored in that chapter. The Ph.D. thesis [Knoblock-91] would probably a good starting point for a more extensive search of that literature.

Finally, some of the ideas explored in this thesis concern fundamental technical issues of probability theory. In §2.11 and chapter 3 I make some proposals which would require a very deep knowledge of the literature in probability and statistics in order to properly evaluate. (For example, a working knowledge of measure theory such as in [Adams-Guillemin-86] would be very useful; see [Good-62].) Unfortunately I am not prepared at present to point to the right areas of related work in the statistics community, since my research into that literature is still very incomplete.

More details about related work will be discussed within each of the individual chapters.