1

Introduction to Practical Reasoning

Practical reasoning of the kind described by philosophers since Aristotle (384–322 BC) is identified as goal-based reasoning that works by finding a sequence of actions that leads toward or reaches an agent's goal. Practical reasoning, as described in this book, is used by an agent to select an action from a set of available alternative actions the agent sees as open in its given circumstances. A practical reasoning agent can be a human or an artificial agent – for example, software, a robot, or an animal. Once the action is selected as the best or most practical means of achieving the goal in the given situation, the agent draws a conclusion that it should go ahead and carry out this action. Such an inference is fallible, as long as the agent's knowledge base is open to new information. It is an important aspect of goal-based practical reasoning that if an agent learns that its circumstances or its goals have changed and a different action might now become the best one available, it can (and perhaps should) "change its mind."

In computer science, practical reasoning is more likely to be known as means-end reasoning (where an end is taken to mean a goal), goal-based reasoning, or goal-directed reasoning (Russell and Norvig, 1995, 259). Practical reasoning is fundamental to artificial intelligence (Reed and Norman, 2003), where it is called means-end analysis (Simon, 1981). In goal-based problem-solving, a search for a solution to a problem is carried out by finding a sequence of actions from available means of solving a problem. An intelligent goal-seeking agent needs to receive information about its external circumstances by means of sensors, and store it in its memory. There are differences of opinion about how practical goal-based reasoning should be modeled. One issue is whether it should be seen as merely an instrumental form of reasoning, or whether it should be also based on values. Many automated systems of practical reasoning for

multi-agent deliberation (Gordon and Richter, 2002; Atkinson et al., 2004a, 2004b; Rahwan and Amgoud, 2006) take values into account.

Chapter 1 introduces the reader to current research on modeling goal-based means-end reasoning and identifies the direction the rest of the book will take in order to build and implement one model. Section 1 explains the simplest forms of practical reasoning. Section 2 explains the two different approaches to practical reasoning in philosophy and artificial intelligence: the belief-desire-intention (BDI) model and the commitment model. Section 3 defines the notion of an autonomous, rational, intelligent agent – the entity, whether human mechanical or animal, that carries out practical reasoning. Section 4 contrasts the practical reasoning of an agent acting alone with that of multi-agent practical reasoning. Sections 5, 6, and 7 show how the simplest form of practical reasoning is the basis of several more complex forms that take into account consideration of alternative actions, chaining of actions, multiple goals, and side effects. These sections show that in addition to instrumental practical reasoning there is also a value-based variant that has proved to be important in computing, especially in artificial intelligence. Section 8 shows how trying to account for all the steps in a complex sequence of practical reasoning results in a state space explosion. Section 9 shows how the various complexities can be accounted for by using an argumentation model with a set of critical questions matching an argumentation scheme for practical reasoning. Section 10, on practical reasoning in multi-agent systems, sums up the findings of Chapter 1 and shows how they lead forward to the work of the subsequent chapters.

1.1 The Basic Form of Practical Reasoning

There are three basic components of a practical inference in the simplest kind of case. One premise describes an agent's goal. A second premise describes an action that the agent could carry out that would be a means to accomplish the goal. The third component is the conclusion of the inference telling us that this action should be carried out. The simplest and most basic kind of practical inference that is readily familiar to all of us can be represented in the following scheme. The first-person pronoun 'I' represents an agent. More correctly, it could be called a rational agent of the kind described by Woodridge (2000), an entity that has goals, some (though possibly incomplete) knowledge of its circumstances, and the capability of acting to alter those circumstances and to perceive (some of) the consequences of so acting.

Basic Form of Practical Inference:
I have a goal, *G*.
Carrying out this action *A* is a means to realize *G*.
Therefore, I ought (practically speaking) to carry out this action *A*.

This basic form of practical inference is very simple, yet we all recognize its importance as a kind of reasoning we use in daily life, and especially in technology of all sorts.

Of course, as we will see, this form of inference is much too simple to represent all of the complications that arise in everyday cases of practical reasoning and that need to be taken into account in any reconstruction of this form of argument. An agent may have many goals, and there may be many ways to carry out a goal. Also, practical reasoning is used in explanations as well as arguments, and thus contexts of use can vary. Still, it is best to start with the simple form of inference given above as representing the most basic kind of practical reasoning. The reason is that this simple form of inference has the most explanatory power. After explaining all the complicating factors that make the basic form too simplistic to work on its own, in the end we will return to something like it.

The basic form of practical inference can be represented as a process of rational deliberation in which an agent reasons from a goal to an action. One way to represent practical reasoning is as a sequence from one state to another in an activity diagram. A diamond-shaped node represents a choice point where an agent needs to make a decision. The basic form of practical reasoning can be represented as a sequence of reasoning in a structure of this kind. The sequence shown in Figure 1.1 has an agent asking at the first choice point whether he has a goal. If not, the sequence of reasoning stops. If so, the agent goes on to ask the question of whether there is a means available to carry out the goal. If not, the sequence of reasoning stops. If so, the agent takes the action that is the means. That is all there is to the procedure. At that point it stops. The activity structure of this simplest form of practical reasoning is displayed in Figure 1.1. In this figure and subsequent activity diagrams, rectangles with rounded corners represent actions the agent needs to carry out. The diamond-shaped node represents a decision point. The black dot means 'start' and the black dot with the white border means 'stop.'

It is easy to see that this most basic form of practical reasoning is too simplistic to represent even the most ordinary kinds of cases that admit of some complexity. The following example of practical inference from Aristotle (*De Motu Animalium* 701 a 18) illustrates the simplistic nature of the basic form. The use of the word 'good' indicates that the reasoning is based on values as well as purely instrumental goals.

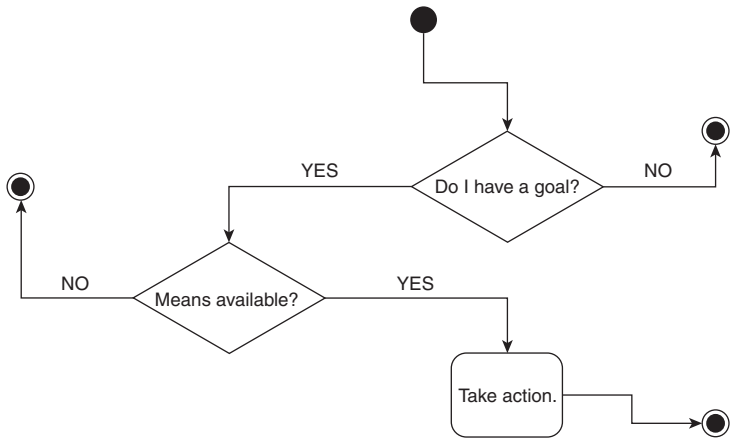

Figure 1.1: The Basic Structure of Simple Practical Reasoning

I should make something good.
A house is something good.
At once I make a house.

The reasoning in this example could be described as "breathless," because the agent immediately leaps to the conclusion to take action. It could even be seen as leaping to a conclusion too quickly, a kind of fallacious reasoning. The fallacy is one of leaping to action immediately without carefully considering the options and costs. This example shows that the basic form, on its own, is inadequate as an argumentation scheme that can be used to represent practical reasoning.

A more complex form of practical reasoning is achieved by chaining together a sequence of practical inferences. The sequence leads toward an ultimate goal, and concludes with a practical directive recommending a course of action to take in a given initial situation, as an agent sees that situation. For example, my goal may be to get to Arnhem this afternoon. The means to get there is for me to take the train that leaves the station at 3:00. The way to get to the train station by 3:00 clock is to take the number 9 bus that leaves the university at 2:30. But in order to take this bus, it may be necessary for me to leave my house by 2:15. It may also be best to get a train ticket. To get the train ticket, I may have to pay some money. Thus, there is a lengthy sequence of actions that I have to carry out in order to fulfill my goal of getting to Arnhem this afternoon. At first, the practical reasoning looks simple, but once I begin to examine it carefully, it breaks down into a complex sequence of connected actions that have to be performed in a particular order. The whole sequence aims at the goal state.

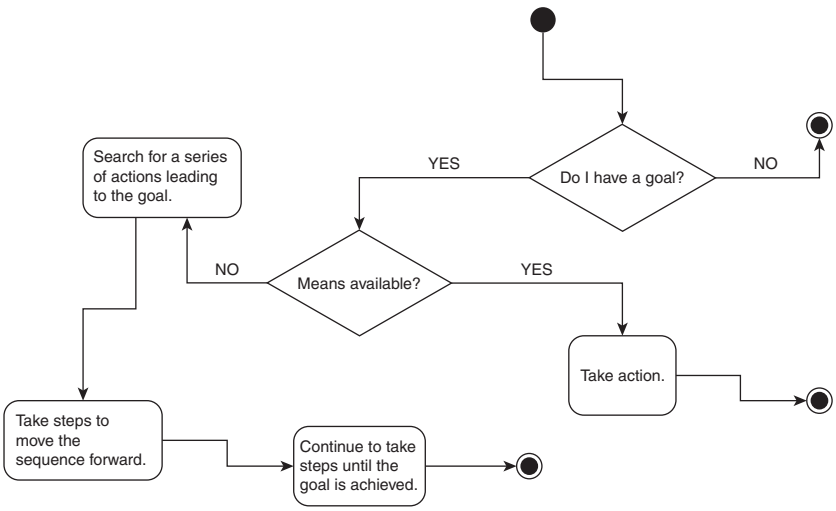

Figure 1.2: Chaining of Practical Inferences.

An example from Aristotle represents a kind of chaining of practical inferences with two steps. The first inference postulates an agent stating a need as a premise, and then finding a means that fulfills the need. In the second step, the agent sees that attaining this means is itself something that requires an action (*De Motu Animalium* 701 a 19).

I need covering.
A cloak is a covering.
Therefore, I need a cloak.
But, what I need I have to make.
And [as concluded above], I need a cloak.
Therefore, I have to make a cloak.

In this case, two practical inferences are joined together to make a chain of practical reasoning. The conclusion of the first inference becomes a premise in the second one.

This kind of case can be generalized. In a sequence of practical reasoning, a series of practical inferences are connected one to the other, as shown in Figure 1.2.

Practical reasoning as represented in Figure 1.2 represents the typical kind of forward chaining that takes place when an agent looks forward in deliberation to try to achieve a goal in the future by carrying out an action now, or before the realization of the future event is contemplated. This could be called the projective use of practical reasoning.

A simple example has been given in Walton (1990a, 89). An agent has the goal of balancing his budget, and decides to try to reduce his heating bill by cutting some firewood with his chainsaw. He thus goes through a complex sequence of actions. He gets the chainsaw and puts it in the trunk of his car. He hitches his trailer to the car. He gets in the car and reaches the woods. Once he has gotten the chainsaw out of the car in the spot he has chosen, he starts it up. The long sequence of how to start the saw is described in the instruction manual. He follows this sequence. He switches on the ignition. He pulls out the choke control. He pushes down the safety catch. He opens the throttle. He puts his foot on the plate behind the rear handle. He grasps the starter handle with the other hand. An ordinary example like this one shows how complex a chain of practical reasoning can be in even the simplest and most everyday case. We all know now that the early attempts in AI to model everyday common sense reasoning were confounded by the realization of how complex the most everyday kind of practical reasoning can be.

In addition to the forward or projective kind of practical reasoning illustrated in the chainsaw case, there is also an abductive or backward use of practical reasoning. In this use, a set of data describing the actions of one agent is known to a second agent or observer. The second agent then concludes to a best explanation of the data, using practical reasoning (Walton, 2004). Typically, the second agent uses a sequence of practical reasoning to connect the first agent's observed actions and presumed goals, putting the whole thing together in what could be called an account. The account serves as the basis of an explanation. There can be several competing explanations that are possible or plausible, and the second agent tries to pick out the best or most plausible one. This is the retrospective or abductive use of practical reasoning. It is very common in law and history. This use of practical reasoning will be studied in Chapter 3.

What has been shown so far is that it is necessary to have more complex forms of practical reasoning in which other relevant considerations are taken into account before drawing a conclusion and closing the case. We now go on to consider these more complex forms of practical reasoning.

1.2 The BDI and Commitment Models

The structure of the practical inference can be brought out in a more useful way if we represent the outcomes of actions as propositions – *A, B, C*, . . . , so that carrying out an action can be described as bringing about a proposition, or "making it true." Using this way of speaking, the structure of the practical

inference can be represented by the following scheme, a variant on the basic form cited above:

My goal is to bring about *A*.
Bringing about *B* is the way to bring about *A*.
Therefore, I should bring about *B*.

It is controversial whether this model represents the right approach. According to one theory of action, an action may be analyzed as the bringing about of an event or state of affairs, something like a proposition that is made true or false by an agent (Horty and Belnap, 1995). As shown in an overview of formal action systems (Segerberg, 1984), some have argued for an opposed approach that sees actions as a species of events.

There are two more specific ways of representing practical inferences, depending on what is meant by the expression 'means' when it is said that an action is a means to achieve a goal. Typically what is referred to is a necessary condition for bringing about something, but in some cases what is referred to is a sufficient condition for bringing about something. For example, paying tuition is a means of graduating, in the sense that it is a necessary – but not a sufficient – condition of graduating. My swatting a mosquito may be a sufficient condition of the mosquito's being dead. But it is not a necessary condition, assuming that the mosquito could have died in some other way.

Audi (1989, 86) recognized the distinction between necessary condition schemes and sufficient conditions schemes for practical reasoning. He offered the following example of a sufficient condition scheme (87).

> I really need a peaceful visit in the country. Accepting their invitation for a weekend in the Catskills would be a good way to have such a visit, so I'll accept it.

The assumption made in this example is that accepting the invitation for a weekend in the Catskills would provide a sufficient condition for having a peaceful visit in the country.

The train to Arnhem example given above, however, suggests that in many common cases of practical reasoning, the means refers to a necessary condition. Von Wright (1963, 161) used the following variant on this sort of example.

X wants to reach the train on time.
Unless X runs he will not reach the train on time.
Therefore, X must run.

In another early essay (1972, 41), von Wright, describing practical reasoning as using the means mentioned in the second premise in order to attain the end mentioned in the first premise, offered this example.

I want to make the hut habitable.
Unless I heat the hut, it will not become habitable.
Therefore, I must heat the hut.

This example, like the train example, suggests that the means being described is meant to be a necessary condition of achieving the goal.

The formulation of the goal premise has been a longstanding controversy. One group of theorists holds that the goal premise should be an expression of the agent's intentions, wants, or desires. For example, Clarke (1985, 17) offered this kind of formulation in which E is an end (goal), M is the means (an action), and C is the set of circumstances beyond the control of the agent.

I want E.
My doing M is a means to attaining E if C obtains.

I should (ought to) do M.

He cited the following example (p. 3): I want to keep dry. Taking this umbrella is a means of keeping dry if it rains. It will rain. I may conclude then that I should take the umbrella. This version, which expresses the major premise as a want, fits the BDI model. Audi (1989, 87) presented a version of the sufficient condition scheme for practical reasoning that also fits the BDI model:

I want this goal.
Carrying out this action is a way for me to bring about this goal under these circumstances.
There is no other way to bring about this goal now, which is as preferable to me, or more preferable to me, than carrying out this action.
There is no sufficient reason for me not to bring about this action under these circumstances.
Therefore, let me carry out this action.

Others have expressed the view that the first premise should express the agent's goal as a commitment. Walton (1990a) based the analysis of practical reasoning on a commitment model. Von Wright, in different places, seemed to accept both models. In his first paper (1963), he used the term "want" predominantly. However, in his 1972 paper he started out using "want" as the key word, but then (1972, 45) described an intention to pursue an end as being "resolved" to go after something in the future. This usage seems more like the language of commitment.

Von Wright did not appear to see practical reasoning as a deductive form of inference, given that he wrote (1972, 59) that the premises do not entail the

behavior stated in the conclusion. Instead, as he put it, the agent is logically bound to his intention "within the teleological frame . . . for his prospective action." But what does that mean? It seems to imply that the binding nature of practical inference is something other than deductive validity. Perhaps it suggests that a prudent agent should adopt a consistent plan of action. But what sort of consistency is that? This question has led to a lot of philosophical theorizing about weakness of will. Suppose the agent has the goal of doing his homework, and has the means to do it, but is just too lazy? Does that mean the agent is inconsistent, in some sense? This problem remains unsolved (Walton, 1997), relating to problems of retraction of commitments in plans.

There are two different philosophical theories about how practical reasoning should be modeled. The commitment-based argumentation approach (Walton and Krabbe, 1995) can be contrasted with the BDI (belief-desire-intention) theory (Bratman, 1987; Bratman, Israel, and Pollack, 1988; Paglieri and Castelfranchi, 2005; Wooldridge, 2002). In argumentation theory, two agents (in the simplest case) interact with each other in a dialogue in which each contributes speech acts. Each has a commitment set, and as the one asks questions that the other answers, commitments are inserted into or retracted from each set, depending on the type of move (speech act) each speaker makes. A commitment is a proposition that an agent has gone on record as accepting. One type of speech act is the putting forward of an argument. When the one agent puts forward an argument, the other can reply, either by asking critical questions or by putting forward a counter-argument.

According to BDI theory, an agent has a set of beliefs that are constantly being altered by sensory input from its environment that continually updates its previous beliefs. From these beliefs, the agent builds up desires (wants) that are then evaluated by desirability and achievability to form intentions. An intention is a persistent goal that is not easily given up. The two models are different because a commitment is not necessarily a belief. Belief implies commitment, but not vice versa. Belief is a psychological notion whereas commitment is a procedural notion based on dialogue rules (Engel, 2000).

An outline of the main features of the BDI model and the thematic variations in the approaches of its chief exponents has been given by Hitchcock (2002). According to Hitchcock, the BDI model was first articulated by Aristotle (*Nicomachean Ethics* III.31112b15-20), who wrote that good deliberation begins with a wish for some end and follows through with a means for attaining it, and with other means that may be needed to carry out the first means.

The conclusion of this process, according to Aristotle, is a decision to take action. In Bratman's (1987) variation of the BDI model, to form an intention to do something is to adopt a plan, and thus intentions, as well as desires (wants) and beliefs, need to be added to the model. Pollock (1995) added what he called "likings," as well as desires, which need to work in combination with beliefs and intentions. Hitchcock argued that although Pollock's system has many advantages, it is incomplete in three important respects:

1. it is solipsistic, in that it does not allow for back-and-forth discussion between agents;
2. it is egoistic, in that it does not take community values or likings into account; and
3. it is unsocial, in that it does not take groups of agents with a governance structure into account.

Searle (2001) also advocated the BDI model of practical reasoning, but, like Bratman, often shifts to the language of commitment. The problem for the BDI model is that it is hard to model practical inference because beliefs are not transferred from the premises to the conclusion of a practical inference. If I believe that proposition *A*, and proposition *B* is a logical consequence of *A*, it need not follow that I believe that *B*. Searle (2001, 241) poses the problem as one of seeking patterns of practical validity such that acceptance of the premises of the valid practical argument commits one to acceptance of the conclusion. However, acceptance – another word for commitment – does not equate with, or necessarily imply, belief. Bratman, who often expresses his view of practical reasoning in terms of commitment (1987, chapter 7), is, however, also seen as an advocate of the BDI model. Bratman, Israel, and Pollack (1988, 347) wrote, "The fundamental observation of our approach is that a rational agent is committed to doing what she plans." Perhaps what is shown is that commitment is associated with planning.

Planning – also called automated planning and scheduling – is a technology used in artificial intelligence, based on an intelligent agent having a set of goals and being able to generate a sequence of actions that leads to the fulfillment of one or more of these goals. Solutions, which often resort to trial and error strategies, are found and evaluated prior to the execution point at which the agent carries out the action. Problem-solving is another area of computing, but problem-solving is similar to planning and both technologies are based on a practical reasoning framework in which an agent concludes to an action based on its goals and what it takes to be the means to lead to fulfillment of a particular goal. Although planning and problem-solving are