Draft Chapter for Handbook on Logic of Normative Systems

Title: The Formalization of Practical Reasoning: Problems and Prospects

Author: Richmond H. Thomason, rthomaso@umich.edu

Version of: March 5, 2012

1. Chapter abstract

Deontic logic, as traditionally conceived, provides only a deductive theory that constrains the states or possible worlds that an agent should try to remain within. As such, it only encompasses a small part of practical reasoning, which is concerned with selecting, committing to, and executing plans of action. In this chapter I try to frame the general challenge that is presented to logical theory by the problem of formalizing practical reasoning, and to survey the existing resources that might contribute to the development of such a formalization. I conclude that, while a robust, adequate logic of practical reasoning is not in place, the materials for developing such a logic are now available.

2. The challenge of formalizing practical reasoning

Practical reasoning is deliberation. It is reasoning about what to do. We do it all the time. Any day in our life will provide us with hundreds of examples of thinking about what to do. But it has been remarkably difficult to produce a comprehensive, adequate theory of practical reasoning. Part of the difficulty is that the topic is studied by different disciplines, each of these has something important to contribute, and it is unusual to find a study of practical reasoning that brings all of these perspectives together.

This section will begin by considering examples of practical reasoning. (I suspect that the range of examples is broader than many people might imagine.) It will then propose a rationale for classifying these examples, and canvass the disciplines that have something useful to say about the reasoning.

This provides materials for thinking about a more comprehensive account of practical reasoning. In the remainder of the paper, I try to say something about what an approach that begins to do justice to the subject might be like.

2.1. Some Examples

All too many published discussions of practical reasoning—even book-length discussions—cover only a very small part of the territory. For that reason, it's vital to begin with a broad range of examples.

Example 1. Ordering a meal at a restaurant.

> The deliberating agent sits down in a restaurant and is offered a menu. Here, the problem is deciding what to eat and drink. Suppose that the only relevant factors are price and preferences about food. Even for a moderately sized menu and wine list, the number of possible combinations is over $400,000$. It would be very unlikely for an ordinary human being to work out a total preference ordering for each option. In fact, even though the decision will probably involve weighing preferences about food and drink against preferences about cost, the reasoning might well produce an order without appealing to a general rule for reconciling these preferences.

Example 2. Deciding what move to make in a chess game.

In chess, an individual action needs to be evaluated in the context of its continuations. There is no uncertainty about the current state or the immediate consequences of actions, but much uncertainty about moves that the opponent might make. The *search space* (i.e., the number of possible continuations) is enormous—on the order of 10^{43}. Determining the value of positions involves conflicting criteria (e.g. positional advantages versus numerical strength); these conflicts must be resolved in comparing the value of different positions. In tournament chess, deliberation time is limited. These somewhat artificial constraints combine to concentrate the reasoning on exploration of a search space. Perhaps because of this, the reasoning involved in chess has been intensively investigated by psychologists and computer scientists, and influenced the classical work on search algorithms in AI; see [Simon and Schaeffer, 1992].

Example 3. Savage's omelet.

In [Savage, 1972][pp. 13–15], Leonard Savage describes the problem as follows.

> Your wife has just broken five good eggs into a bowl when you come in and volunteer to finish making the omelet. A sixth egg, which for some reason must either be used for the omelet or wasted altogether, lies beside the bowl. You must decide what to do with the unbroken egg. ... you must decide between three acts only, namely, to break it into the bowl containing the other five, to break it into a saucer for inspection, or to throw it away without inspection.

This problem involves preferences about the desired outcomes, as well as risk, in the form of a positive probability that the egg is spoiled. The problem is to infer preferences over actions. The outcomes are manifest and involve only a few variables, the preferences over them are evident, and the probabilities associating each action with an outcome can be easily estimated. In this case, the reasoning reduces to the calculation of an expected utility.

Example 4. Designing a house.

This example is less obviously practical; it is possible for an architect to design a house without thinking much about the actions that will go into building it, leaving this to the contractor.[1] However, an architect's design becomes the builder's goals, and I would maintain that inferring goals is a form of practical reasoning. The reasoning combines constraint satisfaction and optimization, where again conflicts between competing desiderata may need to be resolved. Any real-life architect will also use *case-based reasoning*, looking in a library of known designs for one that is relevant, and modifying a chosen example to suit the present purpose.

Example 5. Deciding how to get to the airport.

[1] Of course, a good design has to take into account how to build a house, in order to make sure that the design is feasible.

This is a planning problem; the agent a has an inventory of actions, knows their preconditions and effects, knows the relevant features of the current state, and has as its goal a state in which a is at the airport. In its simplest form, the problem is to find a sequence of actions that will transform the current state into a state that satisfies the goal. Planning, or means-end reasoning, is one of the most intensively studied forms of reasoning in AI. The earliest planning algorithms made many simplifying assumptions about the planning situation and the conditions that a satisfactory plan must meet; over the years, sophisticated planning algorithms have been developed that depend on fewer of these assumptions and so can be used in a variety of realistic settings.[2]

Example 6. Cracking an egg into a bowl.
This is a case in which most of of us do the action automatically, with hardly any conscious reasoning. Probably most people can't remember the circumstances under which they learned how to do it. But the activity is complex there are many ways to get it wrong. This example was proposed as a benchmark problem in the formalization of common-sense reasoning. The literature on this problem shows that the reasoning is complex, and it presupposes much common-sense knowledge; see, for instance, [Shanahan, 1997a]. This example is different from the previous ones in that the solution to the reasoning problem is acted out; the reasoning must engage motor systems, and it depends on these systems for grasping and manipulating objects according to plan. For obvious reasons, Savage ignored this part of the omelet problem.

Example 7. Playing table tennis.
Unlike chess, table tennis is a game in which practical reasoning has to be *online*; engaged in complex, real-time activities involving the perceptual and motor systems. For an novice, the reasoning may be consumed by the need to keep the ball in play; experts may be able to engage in tactical reasoning. But there is no time to spare for reflection; the reasoning needs to be thoroughly connected to the ongoing process of play.

Example 8. Playing soccer.
Soccer is like table tennis, but with the added dimension of teamwork and the need to recognize and execute play. This task was selected as a benchmark problem in robotics, and has been extensively studied. See, for instance, [Visser and Burkhard, 2007, Ros *et al.*, 2009, Asada *et al.*, 1999].

Example 9. Typing a message.

[2]See, for instance, [Reiter, 2001]. For the airport problem in particular, see [Lifschitz *et al.*, 2000].

Typing an email message, composing it as you go along, starts perhaps with a general idea of what to say. The reasoning that produced a rough idea of the content may have taken place reflectively, but once composition has begun, several reasoning processes are engaged simultaneously, and have to be coordinated. The general idea of what to say has to be packaged in linguistic form, and this form has to be rendered by motor actions at the keyboard. For a skilled typist composing a straightforward message, these complex, practical tasks are combined and executed very quickly, perhaps at the rate of 70 words per minute. for this to happen, the interface between high-level linguistic reasoning and motor skills has to be very robust.

Example 10. Factory scheduling.

The factory scheduler has to produce, say on a daily basis, a sequence of manufacturing operations for each order to be processed that day, and a schedule allocating times and machines to these operations. This problem is notorious for the difficulty of the reasoning; it involves horrible combinatorics, uncertainty, limited time for reflection, and the resolution of many conflicting desiderata. Among the goals cited by [Fox and Clarke, 1991] are (1) meeting order dates, (2) minimizing work-in-process time, (3) maximizing allocation of factory resources, and (4) minimizing disruption of shop activity.

Part of the interest of this example lies in the difference in scale between this problem and Savage's omelet problem. It is not clear that there is any way to construct a single, coherent utility function for the task, by reconciling the four desiderata mentioned above. Any reconciliation will leave some managers unhappy: salesmen will favor goal (1), and production managers will favor goals (2)-(3), perhaps giving different weights to these. Nor is it easy to produce a global probability function for a system with so many interacting variables.

Example 11. Ordering dessert.

Let's return the restaurant of Example 1. The main course is over, and our agent is offered a dessert menu and the choice of whether to order dessert. On the one hand, there is a direct desire for dessert, perhaps even a craving. This alternative is colored with and motivated by emotion, even if the emotion is not overwhelming. But suppose that there is a contrary emotion. The agent is unhappy with being overweight and has determined to eat less, and may have told others at the table about the decision to undertake a diet. This creates a conflict, coloring the choice of dessert with negative associations, perhaps even shame. The chief difference between this conflict and those in Examples 2 and 4 is that this decision is emotionally "warm;" the outcome may be influenced by a craving and the presence of the desired object. (Perhaps this is why some restaurants invest in dessert trays.)

Example 12. An unanticipated elevator.

A man decides to visit his stockbroker in person, something he has never done. He takes a bus to a stop near the stockbroker's downtown address, gets off the bus, locates the building and enters it. He finds a bank of elevators, and sees that the stockbroker is on the 22nd floor. This man has a strong dislike for elevators, and is not feeling particularly energetic that day. He reconsiders his plan.

Example 13. A woman is working in her garden.

She becomes hot and tired, and decides to take a break. Or she hears the telephone ringing in her house, and decides to answer it. Or she sees smoke coming out of the window of her house, and runs for help.

Example 14. The wrath of Achilles.

In Book I of *The Iliad*, the hero Achilles is outraged and dishonored by his warlord Agamemnon, who insults him and declares that he will take back, in compensation for his own loss and Achilles' disrespectful behavior, the captive woman that Achilles had received as his war prize.

Homer goes on to describe Achilles' reaction. Achilles is headstrong, but his reaction is partly physical and partly intellectual: his heart pounds with rage, but instead of acting immediately he asks himself a a question: should he draw his sword and kill the king? To explain his decision, the poet brings in a god: Athena, invisible to everyone else, seizes him by the hair and persuades him to give in and be patient.

For our purposes, we can suppose that Athena is a literary device. The outrage leads to a direct desire to kill, but instead of acting on it, Achilles realizes that it would be better to restrain himself.

Even though it is strongly informed by emotion, "hot," reasoning intervenes here between the circumstances that color the alternatives and an ensuing resolution to act.

Example 15. Deciding what to say at a given point in a conversation.

Conversation provides many good examples of deliberative reasoning. Where there is conscious deliberation, it is likely to be devoted to content selection. But the reasoning that goes into deciding how to express a given content can be quite complex.

Certainly, any adequate theory of practical reasoning must at least be compatible with this broad range of cases. Better, it should be capable of saying something about the reasoning involved in all of them. Even better, there should be a single architecture for practical reasoning, capable of dealing with the entire range of reasoning phenomena.[3] No doubt, there are special-purpose cognitive modules (e.g., for managing perception, motor behavior, and some aspects of language). But it would be perverse to formulate a theory of a special type of practical reasoning, such as preference generation, probability estimation, or means-end reasoning, and to postulate a "cognitive module" that performs just this reasoning. This methodology would be likely to produce an *ad hoc* and piecemeal account of practical reasoning.

2.2. Towards a classification

The examples in the previous section suggest a set of features that can be used to classify specimens of deliberative reasoning.

[3]For the idea of a cognitive architecture, see [Newell, 1992].

1. Are only a few variables (e.g., desiderata, causal factors, initial conditions) involved in the decision?
2. Do conflicting preferences need to be resolved in making the decision?
3. Is the time available for deliberation small compared to the time needed for adequate reflection?
4. Is the deliberation immediate? That is, will the intentions that result from the deliberation be carried out immediately, or postponed for future execution?
5. Is the deliberation carried out in "real time" as part of an ongoing activity involving sensory and motor activities?
6. Does the reasoning have to interface closely with sensory and motor systems?
7. Is the activity part of a group or team?
8. Does the context provide a definite, relatively small set of actions, or is the set of actions open-ended?
9. Is there certainty about the objective factors that bear on the decision?
10. Is the associated risk small or great?
11. Is the goal of deliberation a single action, or a sequence of actions?
12. Is continuous time involved?
13. Is the deliberation colored with emotions?
14. Is the action habitual, or automatic and unreflective?
15. Is there conscious deliberation?
16. Are there existing plans in play to which the agent is committed or that already are in execution?

Many of the differences marked by these features are matters of degree, so that the boundaries between the types of reasoning that they demarcate are fluid. This strengthens the case for a general approach to the reasoning. There is nothing wrong with concentrating on a special case to see what can be learned from it. Chess and decision problems that, like Savage's omelet, involve a solution to the "small worlds problem"[4] provide good examples of cases where this methodology has paid off. But to concentrate on these cases without paying any attention to the broad spectrum of examples runs the risk of producing a theory that will not be contribute usefully to something more general.

2.3. Disciplines and approaches

Many different disciplines have something to say about practical reasoning. The main theoretical approaches belong to one of the five following areas.

1. Philosophy
2. Logic
3. Psychology
4. Decision Theory and Game Theory
5. Artificial Intelligence

Of course, there is a good deal of overlap and mixing of these approaches: AI, for instance, is especially eclectic and has borrowed heavily from each of the other fields. But work in each area is colored by the typical problems and methods of the discipline, and—typically,

[4]This is the problem of framing a decision problem, concentrating only on the factors that are relevant.

at least— has a distinctive perspective that is inherited from the parent discipline.

The following discussion of these five approaches is primarily interested in what each has to contribute to the prospects for formalizing practical reasoning.

2.3.1. Philosophy

The topic of practical reasoning goes back to Aristotle. In the Twentieth Century there was a brief revival of philosophical interest in the topic of "practical inference." This coincided more or less with early work on deontic and imperative logic, and was carried out by a group of logically minded philosophers and a smaller group of philosophicaly minded logicians. It is a little difficult to distinguish philosophy from logic in this work; I will more or less arbitrarily classify Kenny and some others as philosophers for the purposes of this exposition, and von Wright as a logician.

Post-Fregean interest in imperative logic seems to have begun about the time of World War 2, with [Jørgensen, 1937-1938, Hofstadter and McKinsey, 1939, Ross, 1941]. Later, in the 1960s,[5] some British philosophers became interested in the topic. This period saw 10 or more articles relevant appearing in journals like *Analysis*. Of these, [Kenny, 1966] seems to have the most interesting things to say about the problem of formalizing practical reasoning.[6]

Kenny begins with Aristotle's practical syllogism, taking several specimens of means-end reasoning from the Aristotelian corpus, and beginning with the following example, based on a passage in *Metaphysics* 1032b19.

Example 16. A doctor prescribing.
This man is to be healed.
If his humors are balanced, he will be healed.
If he is heated, his humors will be balanced.
If he is rubbed, he will be heated.
So I'll rub him.

The premisses of the reasoning, according to Kenny, are either (i) desires or duties, or (ii) relevant facts. And he characterizes the conclusion as an action.[7] Kenny points out that this sort of reasoning doesn't fit Aristotelian syllogistic, and that a straightforward modern formalization of it would be invalid. To put it crudely, the inference from P, $Q \to P$, $R \to Q$, and $S \to R$ to S is invalid.

Here, I think Kenny has indicated an important form of practical reasoning, and pointed out a glaring problem with the propositional calculus as a formalization medium. Unfortunately, the theory that he proposes in this paper doesn't seem to solve the problem of providing an account of validity that matches the reasoning. In fact, there are many glaring problems with the crude Propositional Calculus formalization of Example 16, involving the deductive formulation of the reasoning as well as the faithfulness of the formalization to

[5] Judging from internal evidence, the work of Richard Hare influenced this episode of interest in the topic. Elizabeth Anscombe [Anscombe, 1958] may also have been an influence, as well as G.H. von Wright.

[6] For more about this period, see [Green, 1997].

[7] The Aristotelian texts make it pretty clear that Aristotle considered the conclusion to be an action. But for our purposes, it would work better to think of the conclusion as an expression of intention. In some circumstances—when the deliberation is concerned with immediate action and the reasoning is sufficiently persuasive, there is no gap between intention and action.

the language of the example. The failure of Kenny's proposal and of similar ones at the time seems to originate in a lack of logical resources that do justice to the problem. The Propositional Calculus is certainly not the right tool, and deduction is certainly not the right characterization of the reasoning. The only idea that was explored at the time was that of providing a logic of "imperative inference." This idea might help with one problem: formalizing the first premiss of Example 16, which does not seem like a straightforward declarative. But it can't begin to address the challenge posed by the invalidity of the argument. Besides, the idea of an imperative logic didn't lead to anything very new, because of another trend that was taking place at about the same time.

This trend, which tried to absorb imperative and practical inference into some sort of modal logic, was also underway in the 1960s. [Lemmon, 1965] provides a logic of imperatives that prefigures the STIT approach of [Belnap, Jr. *et al.*, 2001], hence a modal approach that brings in the idea of causing a state of affairs. And [Chellas, 1969], recommends and develops a reduction of imperative logic to a more standard deontic logic. This idea provides formal systems with excellent logical properties. But it does so at the expense of changing the subject, and leaving the central problem unsolved. Reasoning in deontic logic is deductive, and if you formalize typical specimens of means-end reasoning like Example 16 in these systems, the formalizations will be invalid.

Even though the literature shows a sustained series of attempts in this period to formalize practical inference, the work didn't lead to anything like a consensus, and produced no sustainable line of logical development. In retrospect, we can identify several assumptions that rendered the formalization project unsustainable:

(1) These philosophers relied too much on deductive inference, with the propositional calculus as a paradigm, and too little on models;
(2) They tended to work with overly simple formal languages;
(3) They didn't bring actions into the formalization explicitly;
(4) They missed the insight that means-end reasoning is more like abduction or heuristic search than deduction.

As we will see in Section 2.3.5, more recent and quite separate developments in computer science have yielded sophisticated logics of means-end reasoning, effectively solving the formalization problem that led to an earlier philosophical impasse in the 1960's and 1970's. The moral seems to be that formalization projects of this sort can involve multiple challenges, and that it can be hard to address these challenges without a body of applications and a community of logicians committed to formalizing the applications and mechanizing the reasoning.

Meanwhile, philosophers seem to have drawn the conclusion that close attention to the reasoning, and searching for formalizations, is not likely to be productive. In the more recent philosophical work on practical reasoning, it is actually quite difficult to find anything that bears on the formalization problem. Almost entirely, the philosophical literature is devoted to topics that might serve to provide philosophical foundations for the theory of practical reasoning—if there were such a theory. Even if, as Elijah Millgram claims in [Millgram, 2001], the driving issue in the philosophy of practical reasoning is to determine which forms of practical reasoning are correct, philosophers seem to pursue this inquiry with informal and very loose ideas of the reasoning itself. In many cases—for instance, the issue of whether intentions cause actions—no formalization of the reasoning is needed for the philosophical purposes. In other cases, however, a formal theory of practical reasoning might help the philosophy, refining some old issues and suggesting new ones.

Even though some philosophers maintain positions that would sharply limit the scope of practical reasoning (reducing it, for instance, to means-end reasoning), I don't know of any explicit, sustained attempt in the philosophical literature to delineate what scope of practical reasoning should be. I don't see how to do this without considering a broad range of examples, as I try to do above in Section 2.1. But in fact, examples of practical reasoning are thin on the ground in the philosophical literature; in [Millgram, 2001], for instance, I counted only 12 examples of practical reasoning in 479 pages—and many of these were brief illustrations of general points.

2.3.2. Logic

There are few departments of logic, and work in logic bearing on practical reasoning tends to be carried out in the context of either Philosophy or Computer Science, and to be influenced by the interests of the parent disciplines. There are, in fact, two separate strands of logical research, one associated with Philosophy and the other with Artificial Intelligence. These have interacted less than one might wish.

Philosophical logic. Georg Henrik von Wright was explicitly interested in practical reasoning, from both a philosophical and a logical standpoint. Most of his writings on the topic are collected together in [von Wright, 1983]; these were published between 1963 and 1982. Like Kenny, von Wright begins with Aristotle's practical syllogism. But he avoids the problem of invalidity by strengthening premisses that introduce ways of achieving something. Von Wright's version of Example 16 would look like this:

> I want to heal this man.
> Unless his humors are balanced, he will not be healed.
> Unless he is heated, his humors will not be balanced.
> Unless he is rubbed, he will not be heated.
> Therefore I must rub him.

By departing from Aristotle's formulation, von Wright makes it easier to formulate the inference in a deontic logic, and to see how the formalization might be valid. At the same time, he is making it more difficult to fit the formalization to naturally occurring reasoning. As in this example, where, for instance, there is surely more than one way to heat the patient, the means that a deliberator chooses in typical means-end reasoning will not be the only way to achieve the end.

This simplification makes it easier for von Wright to propose modal logic, and in particular deontic logic, as the formalization medium for practical reasoning. Von Wright also characterizes his version of deontic logic as a "logic of action," but all this seems to mean is that the atomic formulas of his language may formalize things of the form 'Agent A does action a.' He has little or nothing to say about reasoning about action.

I will not say much here about the subsequent history of deontic logic as a part of philosophical logic. As the field developed, it acquired its own problems and issues (such as the problem of reparational obligations), but as the philosophers concentrated on declarative formalisms and deductive logic, the relevance to practical reasoning, and even means-end reasoning, that von Wright saw in his in early papers such as [von Wright, 1963], attenuated.

Although the subsequent history of deontic logic was less directly concerned with practical reasoning, it shows a healthy tendency to concentrate on naturally occurring problems that arise in reasoning about obligation. This work, which certainly will be well documented in the present volume, has a place in any general theory of practical reasoning. Obligations play a role as constraints on means-end reasoning, and reasoning about obligations has to be flexible to cope with changing circumstances.

Also, the problem of modeling conditional obligations has produced a large literature on the relationship between modal logic and preference.[8] Of course, reasoning about preferences intrudes into practical reasoning in many ways. How to fit it in is something I am not very clear about at the moment; part of the problem is that so many different fields study preferences, and preferences crop in so many different types of practical reasoning. Maybe the best thing would be to incorporate preferences in a piecemeal way, and hope that a more general and coherent approach might emerge from the pieces.

The Stit approach to agency was already mentioned in Section 2.3.1. This provides a model-theoretic account of how actions are related to consequences that is quite different from the ones that emerged from the attempts in AI to formalize planning. The connections of Stit theory to practical reasoning are tenuous, and I will not have much to say about it.

Philosophy and philosophical logic have served over the years as a source of ideas for extending the applications of logic, and developing logics that are appropriate for the extensions. One would hope that philosophy would continue to play this role. But—at least, for areas of logic bearing on practical reasoning—the momentum has shifted to computer science, and especially to logicist AI and knowledge representation. This trend began around 1980, and has accelerated since. Because many talented logicians were attracted to computer science, and because the need to relate theories to working implementations provided motivation and guidance of a new kind, this change of venue was accompanied by dramatic logical developments, and improved insights into how logic fits into the broader picture. I would very much like to see philosophy continue to play its foundational and creative role in developing new applications of logic, but I don't see how this can happen in the area of practical reasoning unless philosophers study and assimilate the recent contributions of computer scientists.

The point is illustrated by [Gabbay and Woods, 2005]. The paper is rare among contemporary papers in urging the potential importance of a logic of practical reasoning, but—in over 100 pages—it is unable to say what a coherent, sustained research program on the topic might be like. It does mention some important ideas, such as taking the agent into account, as well as nonmonotonic and abductive reasoning, but offers no explicit, articulated theories and in fact is hesitant as to whether logic has a useful role to play, repeating some doubts on this point that have been expressed by some roboticists and cognitive psychologists. Although it cites a few papers from the AI literature, the citations are incidental; work on agent architectures, abductive reasoning, and means-end reasoning goes unnoticed. Part of the problem is that the authors seem to feel that work in "informal logic" might be useful in

[8] See, for instance, [Hansson, 2001, Jones and Carmo, 2002].

approaching the problem of practical reasoning—but the ideas of informal logic are too weak to provide any helpful guidance. If we are interested in accounting for the practical reasoning of agents, we have to include computer programs. For this, we need formal logic—but formal logic that is applicable.

I couldn't agree more with Gabbay and Woods that logicians should be concerned with practical reasoning. But to make progress in this area, we need to build on the accomplishments of the formal AI community.

2.3.3. Psychology

From the beginning of cognitive psychology, a great deal of labor has gone into collecting protocols from subjects directly engaged in problem-solving, much of it practical. Herbert Simon and Allen Newell were early and peristent practitioners of this methodology. This material contains many useful examples; in fact, it helped to inspire early characterizations of means-end reasoning in Artificial Intelligence.

As early as 1947, in [Simon, 1947], Simon had noted divergences between decision-making in organizations and the demands of ideal rationality that are incorporated in decision theory; he elaborated the point in later work. An important later trend that began in psychology, with the work of Amos Tversky and Daniel Kahneman, studies these differences in more detail, providing many generalizations about the way people in fact make decisions and some theoretical models; see, for instance, [Kahneman and Tversky, 1979, Tversky and Kahneman, 1981].

Tversky and Kahneman's experimental results turned up divergences between ideal and actual choice-making that were not obviously due, as Simon had suggested, merely to the application of limited cognitive resources to complex, time-constrained problems. Since their pioneering work, this has become a theme in later research.

All this raises a challenging foundational problem, one that philosophers might be able to help with, if they gave it serious attention. What level of idealization is appropriate in a theory of deliberation? What is the role of "rationality" in this sort of idealization? Is there a unique sort of rationality for all practically deliberating agents, or are there many equally reasonable ways of deliberation, depending on the cognitive organization and deliberative style of the agent? Is the notion of rationality of any use at all, outside the range of a very limited and highly idealized set of decision problems? Probably it would be unwise to address these problems before attempting to provide a more adequate formalization of practical reasoning—that would be likely to delay work on the formalization indefinitely. But the problems are there.

Nowadays, the cognitive psychology of decision-making has migrated into Economics and Management Science, and is more likely to be found in economics departments and schools of business than in psychology departments. This doesn't affect the research methods much, but it does improve the lines of communication between researchers in behavioral economics and core areas of economics. As a result, economic theorists are becoming more willing to entertain alternatives to the traditional theories.

2.3.4. Decision Theory and Game Theory

The literature in these areas, of course is enormous, and most of it has to do with practical reasoning. But traditional work in game theory and decision theory concentrates on problems that can be formulated in an idealized form—a form in which the reasoning can be reduced to deriving an optimum result by calculation.[9] As a result, work in this tradition tends to neglect much of the reasoning in practical reasoning. Of course, an agent must reason to wrestle a practical problem into the required form—to solve Savage's "small worlds problem"—but the literature in economics tends to assume that somehow the problem has been framed, without saying much if anything about the reasoning that might have gone into this process. (Work in decision analysis, of course, is the exception.) And once a problem has been stated in a form that can be solved by calculation, there is little point in talking about deliberative processes.

If we are concerned with the entire range of examples presented in Section 2.1, however, we find many naturally occurring problems that don't fit this pattern; and some of these, at least, exhibit discursive, inferential reasoning. This is one reason why I believe that a general theory of practical reasoning will reserve an important place for qualitative reasoning, and especially for inferential reasoning—the sort of reasoning that gives formalization and logic a foothold. In this respect, Aristotle was on the right track.

At the very least, practical reasoning can involve inference and heuristic search, as well as calculation. (Calculation, of course, is a form of reasoning, but is not inferential, in the sense that I intend.) Any theory of practical reasoning that emphasizes one sort of reasoning at the expense of others must sacrifice generality, confining itself to only a small part of the territory that needs to be covered by an adequate approach. The imperialism of some of those (mainly philosophers, these days) who believe that there is nothing to rationality or practical reasoning other than calculations involving probability and utility, can partly be excused by the scarcity of theoretical alternatives. I will argue in this chapter that the field of Artificial Intelligence has provided the materials for developing such alternatives.

As I said in Section 2.3.3, research in behavioral economics has made microeconomists generally aware that, in their original and extreme form, the idealizations of decision theory don't account well for a broad range of naturally occuring instances of practical reasoning. Attempts to mechanize decision-making led computer scientists to much the same conclusion.

A natural way to address this problem begins with decision theory in its classical form and attempts to relax the idealizations. Simon made some early suggestions along these lines; other, quite different proposals. can be found in [Weirich, 2004] and [Russell and Wefald, 1991]. And other relaxations of decision theory have emerged in Artificial Intelligence: see the discussion of Conditional Preference Nets below, in Section 2.3.5. Still other relaxations have emerged out of behavioral economics, such as Tversky and Kahneman's Prospect Theory;

[9]Microeconomists and statisticians are not the only ones who have taken this quantitative, calculational paradigm to heart. Many philosophers have accepted the paradigm as a model of practical reasoning and rationality. See, for instance, [Skyrms, 1990], a book-length study of practical deliberation, which takes the only relevant theoretical paradigms to be decision theory and game theory, and takes them pretty much in the classical form. Skyrms' book and the many other philosophical studies along these lines have useful things to say; my only problem with this literature is the pervasive assumption that practical reasoning can be comprehensively explained by quantitative theories based on the assumption that agents have global probability and utility functions.

see [Kahneman and Tversky, 1979].

Programs of this sort are perfectly compatible with what I will propose here. A general account of practical reasoning has to include calculations that somehow combine probability (represented somehow) and utility (represented somehow), in order to estimate risk. The more adaptable these methods of calculation are to a broad range of realistic cases, the better. I do want to insist, however, that projects along these lines can only be part of the story. Anyone who has monitored their own decision making must be aware that not all practical reasoning is a matter of numerical calculation; some of it is discursive and inferential. A theory that does justice to practical reasoning has to include both forms of reasoning. From this point of view, the trends from within economics that aim at practicalizing game theory and decision theory are good news. From another direction, work in Artificial Intelligence that seeks to incorporate decision theory and game theory into means-end reasoning is equally good news.[10]

In many cases of practical reasoning, conflicts need to be identified and removed or resolved. Work by economists on value tradeoffs is relevant and useful here; the classical reference is [Keeney and Raiffa, 1976], which contains analyses of many naturally occuring examples.

2.3.5. Computer science and artificial intelligence

For most of its existence, the field of AI has been concerned with realistic decision problems, and compelled to formalize them. As the field matured, the AI community looked beyond procedural formalizations in the form of programs to declarative formulations and logical theories. Often AI researchers have had to create their own logics for this purpose.[11] Here, I will be concerned with three trends in this work: those that I think have most to offer to the formalization of practical reasoning. These are means-end reasoning, reasoning about preferences, and agent architectures.

Dynamic logic and imperative inference. When an agent is given instructions and intends to carry them out unquestioningly, there is still reasoning to be done, and the reasoning is practical[12]—although, as the instructions become more explicit, the less scope there is for interesting reasoning from the human standpoint. Even so, the case of computer programs, where explicitness has to be carried out ruthlessly, can be instructive, because it shows how logical theory can be useful, even when the reasoning paradigm is not deductive.

A computer program is a (possibly very large and complex) imperative, a detailed instruction for carrying out a task. Many of its components, such as

$$\textbf{let } y \textbf{ be } x$$

[10]For a survey, now getting rather old, see [Blythe, 1999]. For an example of a more recent, more technical paper, see [Sanner and Boutilier, 2009].

[11]All along, John McCarthy has been a strong advocate of this approach, and has done much of the most important work himself. See [McCarthy and Hayes, 1969] for an early statement of the methodology, and a highly influential proposal about how to formalize means-end reasoning.

[12]See [Lewis, 1979].

("set the value of y to the current value of x") are imperatives, although some components, like the antecedent of the conditional instruction

$$\textbf{if } (x < y \textbf{ and } \textbf{not}(x = 0)) \textbf{ then let } z \textbf{ be } y/x$$

are declarative.

Inference, in the form of proofs or a model theoretic logical consequence relation, plays a small part in the theory of dynamic logic. Instead, *execution* is crucial. This idea is realized as the series of states that the agent (an idealized computer) goes though when, starting in a given initial state, it executes a program. Because states can be identified with assignments to variables, there are close connections to the familiar semantics of first-order logic.

Dynamic logic is useful because of its connection to *program verification*. A program specification is a condition on what state the agent will reach if it executes the program; if the initial state of a parsing program for an English grammar G, for instance, describes a string of English words, the program execution should eventually halt. Furthermore, (1) if the string is grammatical according to G, the executor should reach a final state that describes a parse of the string, and (2) if the string is not grammatical according to G it should reach a final state that records its ungrammaticality.

Dynamic logic has led to useful applications and has made important and influential contributions to logical theory. It is instructive to compare this to the relatively sterile philosophical debate concerning "imperative inference" that took place in the 1960s and early 1970s.[13] To a certain extent, the interests of the philosophers who debated imperative inference and the logicians who developed dynamic logic were different. Among other things, the philosophers were interested in applications to metaethics, and computational applications and examples didn't occur to them.

But the differences between philosophers and theoretical computer scientists, I think are relatively unimportant; some of the philosophers involved in the earlier debate were good logicians, and would have recognized a worthwhile logical project if it had occurred to them. In retrospect, three factors seem to have rendered the earlier debate unproductive:

1) Too great a reliance on deductive paradigms of reasoning;
2) Leaving a model of the executing agent out of the theoretical picture;
3) Confining attention to simple examples.

In dynamic logic, the crucial semantic notion is the correctness of an imperative with respect to a specification. Logically interesting examples of correctness are not likely to present themselves without a formalized language that allows complex imperatives to be constructed, and without examples of imperatives that are more complicated than 'Close the door'. (The first example that is presented in [Harel *et al.*, 2000] is a program for computing the greatest common divisor of two integers; the program uses a *while*-loop.) And, of course, a model of the executing agent is essential to the logical theory. In fact, what is surprising is how much logic can be accomplished with such a simple and logically conservative agent model.

As I said, the activity of interpreting and slavishly executing totally explicit instructions is a pretty trivial form of practical reasoning. But a logic of this activity is at least a start.

[13]See, for instance, [Williams, 1963, Geach, 1963] as well as [Kenny, 1966], which was discussed above, in Section 2.3.1.

I want to suggest that, in seeking to formalize practical reasoning, we should be mindful of these reasons for the success of dynamic logic, seeking to preserve and develop them as we investigate more complex forms of practical reasoning.

Planning and the formalization of means-end reasoning. Perhaps the most important contribution of AI to practical reasoning is the formalization of means-end reasoning, along with appropriate logics, and an impressive body of research into the metamathematical properties of these logics, and their implementation in planning systems.[14]

This approach to means-ends reasoning sees a planning problem as consisting of the following components:

(1) An initial state. (This might be described by a set of literals, or positive and negative atomic formulas.)

(2) Desiderata or goals. (These might consist of a set of formulas with one free variable; a state that satisfies these formulas is a goal state.)

(3) A set of actions or operators. Each action a is associated with a causal axiom, saying that if a state s satisfies certain preconditions, then a state $\textsc{Result}(a, s)$ that results from performing a in s will satisfy certain postconditions.

Here, the fundamental logical problem is how to define the state or set of successor states[15] resulting from the performance of an action in a state. (Clearly, not all states satisfying the postconditions of the action will qualify, since many truths will carry over to the result by "causal inertia.") This large and challenging problem spawned a number of subproblems, of which the best-known (and most widely misunderstood) is the *frame problem*. Although no single theory has emerged from years of work on this problem as a clear winner, the ones that have survived are highly sophisticated formalisms that not only give intuitively correct results over a wide range of test cases, but provide useful insights into reasoning about actions. Especially when generalized to take into account more realistic circumstances, such as uncertainty about the current state and concurrency or nondeterminism, these planning formalisms deliver logical treatments of means-end reasoning that go quite far towards solving the formalization problem for this part of practical reasoning.

I will try to say more about how these developments might contribute to the general problem of formalizing practical reasoning below, in Section 3.3.

Reasoning about preferences It is hard to find AI applications that don't involve making choices. In many cases, it's important to align these choices with the designer's or a user's preferences. Implementing such preference-informed choices requires (i) a representation framework for preferences, (ii) an elicitation method that yields a rich enough body of preferences to guide the choices that need to be made, and (iii) a way of incorporating the preferences into the original algorithm.

[14][Allen *et al.*, 1990] is a collection of early papers in the field. Both [Shanahan, 1997b] and [Reiter, 2001] describe the earlier logical frameworks and their later generalizations; [Reiter, 2001] also discusses implementation issues.

[15]Depending on whether we are working with the deterministic or the nondeterministic case.

Any attempt to extract the utilities needed for even a moderately complex, realistic decision problem will provide motives for relaxing the classical economic models of utility; but the need for workable algorithms seems to sharpen these motives. See [Goldsmith and Junker, 2008] for examples and details, and [Doyle, 2004], which provides a wide-ranging foundational discussion of the issues, with many references to the economics literature.

Of the relaxations of preference that have emerged in AI, *Ceteris Paribus* Preference Nets are one of the most widely used formalisms.[16] As in multi-attribute utility theory, the outcomes to be evaluated are characterized by a set of features. A parent-child relation must be elicited from a human subject; this produces a graph called a *CP-net*. The parents of a child feature are the features that directly influence preferences about the child. For instance, the price of wheat in the fall (high or low) might influence a farmer's preferences about whether to plant wheat in the spring. If the price will be high, the farmer prefers to plant wheat; otherwise, he prefers not to plant it. On the other hand, suppose that in the farmer's CP-net the price of lumber is unrelated to planting wheat. It can then be assumed that preferences about planting wheat are independent of the price of lumber.

To complete the CP-net, a preference ranking over the values of a child feature must be elicited for each assignment of values to each of the parent features.

Acyclic CP-nets support a variety of reasoning applications (including optimization), and—combined with means-end reasoning—provide an approach to preference-based planning.[17] And in many realistic cases it is possible to extract the information needed to construct a CP-net.

There are extensions of this formalism that allow for a limited amount of reasoning about the priorities of features in determining overall preferences; see [Brafman *et al.*, 2006].

The work in AI on preferences, like decision analysis, tends to concentrate on extracting preferences from a user or customer. Thinking about practical reasoning, however, produces a different emphasis. Some of the examples in Section 2.1—for instance, Examples 1, 4, 10, and 11—were designed to show that preferences are not automatically produced by the environment, by other agents, by the emotions, or by a combination of these things. We deliberate about what is better than what, and preferences can be the outcome of practical reasoning.[18] The status of an agent trying to work out its own preferences, and of a systems designer or decision analyst trying to work out the preferences of a person or an organization, may be similar in some ways, but I don't think we can hope that they are entirely the same. Nevertheless, insights into methods for extracting preferences from others might be helpful in thinking about how we extract our own preferences.

Agent architectures. A nonexecuting planning agent is given high-level goals by a user, as well as the declarative information about actions and the current state of things, as well perhaps as preferences to be applied to the planning process. With this information, it performs means-end reasoning and passes the result along to the user in the form of a plan.

This agent is not so different from the simple instruction-following agent postulated by dynamic logic; its capabilities are limited to the execution of a planning program, and it has

[16] See, for instance, [Domschlak, 2002, Boutilier *et al.*, 2003].

[17] See [Baier and McIlraith, 2008] for details and further references.

[18] For some preliminary and sketchy thoughts about this, see [Thomason, 2002].

little or no autonomy. But—especially in time-limited planning tasks—it may be difficult to formulate a specification, because the notion of what counts as an optimal plan in these condition is unclear.

When the planning agent is equipped with means of gathering its own information, perhaps by means of sensors, and is capable of performing its own actions, the situation is more complicated, and more interesting. Now the agent is interacting directly with its environment, and not only produces a plan, but must adopt it and put it in to action. This has a number of important consequences. The agent will need to perform a variety of cognitive functions, and to interleave cognitive performances with actions and experiences.

(1) Many of the agent's original goals may be conditional, and these goals may be activated by new information received from sensors. This is not full autonomy, but it does provide for new goals that do not come from a second party.
(2) Some of these new goals may be urgent; so the agent will need to be interruptable.
(3) It must commit to plans—that is, it must form intentions. These intentions will constrain subsequent means-end reasoning, since conflicts between its intentions and new plans will need to be identified and eliminated.
(4) It will need to schedule the plans to which it has committed.
(5) It will need to monitor the execution of its plans, to identify flaws and obstacles, and repair them.

Recognizing such needs, some members of the AI community turned their attention from inactive planners to *agent architectures*, capable of integrating some of these functions. Early and influential work on agent architectures was presented in [Bratman *et al.*, 1988]; this work stressed the importance of intentions, and the role that they play in constraining future planning.

Any means-end reasoner needs desires (in the form of goals) and beliefs (about the state of the world and the consequences of actions). As Bratman, Israel, and Pollock point out, an agent that is implementing its own plans also needs to have intentions. Because of the importance of these three attitudes in the work that was influenced by these ideas, architectures of this sort are often known as *BDI architectures*. For an extended discussion of BDI architectures, with references to the literature up to 2000, see [Wooldridge, 2000]. See also [Georgeff *et al.*, 1999].

Work in "cognitive robotics" provides a closely related, but somewhat different approach to agent architectures. Ray Reiter, a leading figure in this area, developed methods for integrating logical analysis with a high-level, programming language called GoLog, an extension of Prolog. Reiter's work is continued by the Cognitive Robotics Group at the University of Toronto.

Developments in philosophical logic and formal semantics have provided logics and models for propositional attitudes; for instance, see [Fagin *et al.*, 1995, Fitting, 2009]. Using these techniques, it is possible to formulate a metatheory for BDI agency. Such a metatheory is not the architecture; the reasoning modules of a BDI agent and overall control of reasoning has to be described procedurally. But the metatheory can provide specifications for some of the important reasoning tasks. Wooldridge's logic of rational agents, $\mathcal{LORA}$, develops this idea; see [Wooldridge, 2000].

A final word. Logicist AI has struggled to maintain a useful relation to applications, in the form of workable technology. Although the struggle has been difficult, many impressive success stories have emerged from this work—enough to convince the larger AI community of the potential value of this approach. The incentive to develop working applications has, I believe, been very helpful for logic, enabling new ideas that would not have been possible without the challenges posed by complex, realistic reasoning tasks.

Practical reasoning is not quite the same as logicist AI, or even the logical theory of BDI agents. But the successful use of logical techniques in this area of AI provides encouragement for a logical approach to practical reasoning. And, of course, it provides a model for how to proceed.

3. Towards a formalization

The challenge is this: how to bring logical techniques to bear on practical reasoning, and how to do this in a way that is illuminating, explanatory, and useful? In this chapter, I will only try to provide an agenda for addressing this challenge. The agenda divides naturally into subprojects. Some of these subprojects can draw on existing work, and especially on work in AI, and we can think of them as well underway or even almost completed. Others are hardly begun.

3.1. Relaxing the demands of formalization

Let's return to the division between theoretical and practical reasoning.

Traditionally, theoretical reasoning domains are formalized using what Alonzo Church called the "logistic method."[19] This method aims to formulate a formal language with an explicit syntax, a model-theoretically characterized consequence relation, and perhaps a proof procedure. Traditional formalizations did not include a model of the reasoning agent, except perhaps, in the highly abstract form of a Turing machine—this sort of agent is guaranteed whenever the consequence relation is recursively enumerable.

When it comes to practical reasoning, I believe that we have to be prepared to relax Church's picture of logical method.[20] My own proposal for a relaxation is this: (1) we need to add a model of the reasoning agent, (2) we need to identify different phases of practical reasoning in agent deliberation, and different ways in which logic might be involved in each phase of the reasoning, and (3) consequently, we need to be prepared to have a logical treatment that is more pluralistic and less unified.

3.2. Agent architectures and division of logical labor

How should we model an agent that is faced with practical reasoning problems? In Section 2.1, I suggested that we should aim at, or at least acknowledge the existence of, a very

[19] [Church, 1959][pp. 47–58].

[20] In fact, writing in 1956, Church was uncomfortable with semantics and model theory. He included these topics, but in a whisper, using small type. Over 50 years later, we have become quite comfortable with model theory and semantics, and are more likely to insist on this ingredient than on proof procedures. And in areas where logic is applied, we have become increasingly comfortable with the idea of bringing the reasoning agent into the picture.

broad range of reasoning problems. Suppose, for instance, that we classify the types of reasoning that we may need to consider in terms of the sort of conclusion that is reached. In view of the examples that were presented in Section 2.1, we will need to be prepared for the agent to infer:

(1) Goals, which then invoke planning processes;
(2) Plans, and the subgoals or means that emerge from plans;
(3) Preferences emerging from reasoning about tradeoffs and risk;
(4) Intentions, commitments about what to do, and (to an extent) about when to do it;
(5) Immediate decisions about what plan to execute;
(6) Immediate, engaged adjustments of ongoing activities and plan executions, and shifts of attention that can affect the task at hand.

The examples in Section 2.1 were chosen, in part, to illustrate these activities. These sorts of deliberation are distinct, and all are practical. Although some of them can be automatic, they all can involve deliberate reasoning.

These six activities comprise my (provisional) division of practical reasoning into subtasks, and of the deliberating agent into subsystems. Each of them provides opportunities for logical analysis and formalization. I will discuss them in turn.

3.3. Means-end reasoning

This is the best developed of the six areas. We can refer to the extensive AI literature on planning and means-end reasoning not only for well developed logical theories, but for ideas about how this deliberative function interacts with the products of other deliberative subsystems—for instance, with preferences, and with plan monitoring and execution.

3.4. The practicalization of desires

On the other hand, work in AI on means-end reasoning, and on BDI agents, has little or nothing to say about the emotions and the origins of desires. In general, it is assumed that these come from a user—although the goals may be conditional, so that they are only activated in the appropriate circumstances. In principle, there is no reason why goals couldn't be inferred or learned. But the relevant reasoning processes have not, as far as I know, been formalized.

In truly autonomous agents some desires—perhaps all—originate in the emotions. Although a great deal has been written about the emotions, it is hard to find work that could find a useful purpose of logic.[21]

However desires originate, although they may be emotionally colored, they may not all be emotionally "hot;" and, to be useful in reasoning, some desires must be conditional, and self-knowledge about conditional desires must be robust. My preference for white wine this evening will probably be colored by feelings of pleasure when I think about the refreshing

[21] Not [Solomon, 1976], which has a chapter on "Reason and the passions," a section on "The Rationality of the emotions," and a chapter on "The logic of the emotions." Not [Minsky, 2006], written by an author who knows something about AI. But work on modeling artificial characters for applications in areas like interactive fiction might be useful; see [Bates, 1994].

taste of white wine. But the feeling of hypothetical pleasure is relatively mild; I am certainly not carried away by the feeling. And AI systems builders are interested in obtaining a large body of conditional preferences from users because preferences need to be brought to bear under many different circumstances, so that a user's unconditional preferences—the preferences that are activated in the actual state of affairs—will not be very useful. Fully autonomous agents need conditional preferences as well, in planning future actions and in contingency planning.

Perhaps—to develop the example of preference for white wine a bit further—the only mechanism that is needed to generate conditional desires is the ability to imagine different circumstances, together with the ability to color these circumstances as pleasant (to some degree), and unpleasant (to some degree). But it is unlikely to be this simple, because pleasantness is not monotonic with respect to information: I find the idea of a glass of white wine quite pleasant, but the idea of a glass of white wine with a dead fly in it quite unpleasant. Also, my feelings about some imagined situations can be mixed, with elements that I find pleasant and elements that I find unpleasant. At this point, I might have to invoke a conflict resolution method that has little or nothing to do with the emotions.

This leads to a further point: there is a difference between raw or immediate desires, or *wishes*, and all-things-considered desires, or *wants*. This is because desires can not only conflict with one another, but with beliefs. And, when they conflict with beliefs, desires must be overridden: to do otherwise would be to indulge in wishful thinking.

In [Thomason, 2000], I explored the possibility of using a nonmonotonic logic to formalize this sort of practicalization of desires. The target reasoning consisted of deliberations such as the following. (The deliberator is a hiker who forgot her rain gear.)

1. I think it's going to rain.
2. If it rains, I'll get wet.
3. If I get wet, I'll stay wet unless I give up and go home.
4. I wouldn't like to stay wet.
5. I wouldn't like to give up and go home.

The argument reaches an impasse, and a conflict needs to be addressed to resolve it. There are two possible conclusions here, depending on how the conflict is resolved:

6. On the whole, I'd rather go home.

6′. On the whole, I'd rather go on hiking.

The main purpose of Steps 1–5 is to identify the conflict.

I'm not altogether happy with the theory presented in [Thomason, 2000], but I still believe that the practicalization of desires is an important part of practical reasoning that provides opportunities for using logic to good advantage.

3.5. Intention formation

The product of successful means-end deliberation will be an intention, taking the form of commitment to a plan. But the deliberation would not get started without a goal—and I see no difference between a goal and a (perhaps very general and sketchy) intention. Often, even in human agents, these goals come from habits, or from compliantly accepted instructions from other agents.

But sometimes goals arise internally, as outcomes of deliberation. The hiker in Section 3.4 provides an example. If the conclusion of the reasoning is a practicalized desire to turn back and head for home, commitment to the conclusion will produce an intention, which may even become a goal for means-end reasoning. ("How am I to get home?")

This is why practicalization can be an important component of practical reasoning, especially if the reasoner is an autonomous human being.

3.6. What to do now?

There will come moments in the life of an autonomous agent when there is scope for new activities. These opportunities need to be recognized, and an appropriate task needs to be selected for immediate execution. A busy agent with many goals and a history of planning may have an agenda of tasks ready for such occasions; but even so, it may take reasoning to select a task that is rewarding and appropriate. I do not know if any useful work has been done on this reasoning problem.

3.7. Scheduling, execution and engagement

Some of the examples in Section 2.1 were intended to illustrate the point that there can be deliberation even in the execution of physically demanding, real-time tasks. And there can be such a thing as overplanning, since the plans that an agent makes and then performs itself will need to be adjusted to circumstances.

Also, not all intentions are immediate. Those that are not immediate need to be invoked when the time and occasion are right.

There has been a great deal of useful work on these topic in AI; just one one recent example is [Fritz, 2009].

3.8. Framing a practical problem

Leonard Savage's "Small worlds problem" is replicated in the more qualitative setting of means-end deliberation. A means-end reasoning problem requires (at least) a set of actions, a description of the initial conditions, and a goal. But, even in complex cases, formulations of planning problems don't include every action an agent might perform, or every fact about the current state of the world. Somehow, a goal (like "getting to the airport") has to suggest a method of distinguishing the features of states (or "fluents") and the actions that are relevant and appropriate.

I'm sure that ontologies would be helpful in addressing this problem, but other than this I have very little to say about it at the moment.

Bibliography

[Allen *et al.*, 1990] James Allen, James Hendler, and Austin Tate, editors. *Readings in Planning.* Morgan Kaufmann, San Mateo, California, 1990.

[Anscombe, 1958] G.E.M. Anscombe. *Intention.* Blackwell Publishers, Oxford, 1958.

[Asada *et al.*, 1999] Minoru Asada, Hiroaki Kitano, Itsuki Noda, and Manuela Veloso. RoboCup today and tomorrow—what we have learned. *Artificial Intelligence*, 110(2):193–214, 1999.

[Baier and McIlraith, 2008] Jorge A. Baier and Sheila A. McIlraith. Planning with preferences. *The AI Magazine*, 29(4):25–36, 2008.

[Bates, 1994] Joseph Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.

[Belnap, Jr. *et al.*, 2001] Nuel D. Belnap, Jr., Michael Perloff, and Ming Xu. *Facing the Future: Agents and Choices in Our Indeterminist World.* Oxford University Press, Oxford, 2001.

[Blythe, 1999] Jim Blythe. An overview of planning under uncertainty. In Michael Wooldridge and Manuela Veloso, editors, *Artificial Intelligence Today*, pages 85–110. Springer-Verlag, Berlin, 1999.

[Boutilier *et al.*, 2003] Craig Boutilier, Ronen I. Brafman, Carmel Domschlak, Holger H. Hoos, and David Poole. CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research*, 21:135–191, 2003.

[Brafman *et al.*, 2006] Ronen I. Brafman, Carmel Domshlak, and Solomon E. Shimony. On graphical modeling of preference and importance. *Journal of Artificial Intelligence Research*, 25:389–424, 2006.

[Bratman *et al.*, 1988] Michael E. Bratman, David Israel, and Martha Pollack. Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4:349–355, 1988.

[Chellas, 1969] Brian Chellas. *The Logical Form of Imperatives.* Perry Lane Press, Stanford, California, 1969.

[Church, 1959] Alonzo Church. *Introduction to Mathematical Logic, Vol. 1.* Princeton University Press, Princeton, 1959.

[Domschlak, 2002] Carmel Domschlak. *Modeling and Reasoning about Inferences with CP-Nets.* Ph.d. dissertation, Ben-Gurion University of the Negev, Be'er Sheva, 2002.

[Doyle, 2004] Jon Doyle. Prospects for preferences. *Computational Intelligence*, 20(2):111–136, 2004.

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

[Fitting, 2009] Melvin Fitting. Intensional logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* The Metaphysics Research Lab, Stanford, California, 2009.

[Fox and Clarke, 1991] J. Fox and M. Clarke. Towards a formalization of arguments in decision making. In *Proceedings of the 1991 AAAI Spring Symposium on Argument and Belief*, pages 92–99. AAAI, 1991.

[Fritz, 2009] Christian Fritz. *Monitoring the Generation and Execution of Optimal Plans*. Ph.d. dissertation, University of Toronto, Toronto, 2009.

[Gabbay and Woods, 2005] Dov M. Gabbay and John Woods. The practical turn in logic. In Dov M. Gabbay and Franz Guenthner, editors, *Handbook of Philosophical Logic, Volume XIII*, pages 1–122. Springer-Verlag, Berlin, 2 edition, 2005.

[Geach, 1963] Peter T. Geach. Imperative inference. *Analysis*, 23, Supplement 1(3):37–42, 1963.

[Georgeff *et al.*, 1999] Michael Georgeff, Barney Pell, Martha Pollack, Miland Tambe, and Michael Wooldridge. The belief-desire-intention model of agency. In Jörg P. Müller, Munidar P. Singh, and Anand S. Rao, editors, *Intelligent Agents V: Agents Theories, Architectures, and Languages*, pages 1–10. Springer-Verlag, Berlin, 1999.

[Goldsmith and Junker, 2008] Judy Goldsmith and Ulrich Junker. Preference handling for artificial intelligence. *The AI Magazine*, 29(4):9–12, 2008.

[Green, 1997] Mitchell Green. The logic of imperatives. In E. Craig, editor, *The Routledge Encyclopedia of Philosophy*, pages 717–21. Routledge, New York, 1997.

[Hansson, 2001] Sven Ove Hansson. Preference logic. In Dov M. Gabbay and Franz Guenthner, editors, *Handbook of Philosophical Logic, Volume IV*, pages 319–394. Kluwer Academic Publishers, Amsterdam, 2 edition, 2001.

[Harel *et al.*, 2000] David Harel, Dexter Kozen, and Jerzy Tiuryn. *Dynamic Logic*. The MIT Press, Cambridge, Massachusetts, 2000.

[Hofstadter and McKinsey, 1939] Albert Hofstadter and J.C.C. McKinsey. On the logic of imperatives. *Philosophy of Science*, 6:446–457, 1939.

[Jones and Carmo, 2002] Andrew J.I. Jones and José Carmo. Deontic logic and contrary-to-duties. In Dov M. Gabbay and Franz Guenthner, editors, *Handbook of Philosophical Logic, Volume VIII*, pages 265–344. Kluwer Academic Publishers, Amsterdam, 2 edition, 2002.

[Jørgensen, 1937-1938] Jørgen Jørgensen. Imperatives and logic. *Erkenntnis*, 7:288–296, 1937-1938.

[Kahneman and Tversky, 1979] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

[Keeney and Raiffa, 1976] Ralph H. Keeney and Howard Raiffa. *Decisions With Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley and Sons, Inc., New York, 1976.

[Kenny, 1966] Anthony J. Kenny. Practical inference. *Analysis*, 26(3):65–75, 1966.

[Lemmon, 1965] Edward John Lemmon. Deontic logic and the logic of imperatives. *Logique et Analyse*, 8:39–71, 1965.

[Lewis, 1979] David K. Lewis. A problem about permission. In Esa Saarinen, Risto Hilpinen, Ilkka Niiniluoto, and Merrill Province Hintikka, editors, *Essays in Honour of Jaakko Hintikka*. D. Reidel Publishing Co., Dordrecht, Holland, 1979.

[Lifschitz *et al.*, 2000] Vladimir Lifschitz, Norman McCain, Emilio Remolina, and Armando Tacchella. Getting to the airport: The oldest planning problem in AI. In Jack Minker, editor, *Logic-Based Artificial Intelligence*, pages 147–165. Kluwer Academic Publishers, Dordrecht, 2000.

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

[Millgram, 2001] Elijah Millgram. Practical reasoning: The current state of play. In Elijah Millgram, editor, *Varieties of Practical Reasoning*, pages 1–26. The MIT Press, Cambridge, Massachusetts, 2001.

[Minsky, 2006] Marvin Minsky. *The Emotion Machine*. Simon & Schuster, New York, 2006.

[Newell, 1992] Allen Newell. *Unified Theories of Cognition*. Harvard University Press, Cambridge, Massachusetts, 1992.

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

[Ros *et al.*, 2009] Raquel Ros, Josep Llus Arcos, Ramon Lopez de Mantaras, and Manuela M. Veloso. A case-based approach for coordinated action selection in robot soccer. *Artificial Intelligence*, 173(9–10):1014–1039, 2009.

[Ross, 1941] Alf Ross. Imperatives and logic. *Theoria*, 7:53–71, 1941. Reprinted with minor changes, in *Philosophy of Science*, Vol. 11 (1944), pp. 30–46.

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

[Sanner and Boutilier, 2009] Scott Sanner and Craig Boutilier. Practical solution techniques for first-order MDPs. *Artificial Intelligence*, 173(5–6):748–788, 2009.

[Savage, 1972] Leonard Savage. *The Foundations of Statistics*. Dover, New York, 2 edition, 1972.

[Shanahan, 1997a] Murray Shanahan. A logical formalisation of Ernie Davis' egg cracking problem. Unpublished manuscript, Imperial College London, 1997.

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

[Simon and Schaeffer, 1992] Herbert A. Simon and Jonathan Schaeffer. The game of chess. In Robert J. Aumann and Sergiu Hart, editors, *Handbook of Game Theory with Economic Applications, Vol. 1*, pages 1–17. North-Holland, Amsterdam, 1992.

[Simon, 1947] Herbert A. Simon. *Administrative Behavior*. The Macmillan Company, New Yori, 1947.

[Skyrms, 1990] Brian Skyrms. *The Dynamics of Rational Deliberation*. Harvard University Press, Cambridge, Massachusetts, 1990.

[Solomon, 1976] Robert C. Solomon. *The Passions: The Myth and Nature of Human Emotion*. Anchor Press, New York, 1976.

[Thomason, 2000] Richmond H. Thomason. Desires and defaults: A framework for planning with inferred goals. In Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, editors, *KR2000: Principles of Knowledge Representation and Reasoning*, pages 702–713, San Francisco, 2000. Morgan Kaufmann.

[Thomason, 2002] Richmond H. Thomason. Preferences as conclusions. In Ulrich Junker, editor, *Preferences in AI and CP: Symbolic Approaches*, pages 94–98. AAAI Press, Menlo Park, California, 2002.

[Tversky and Kahneman, 1981] Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science, New Series*, 211(4481):453–458, 1981.

[Visser and Burkhard, 2007] Ubbo Visser and Hans-Dieter Burkhard. RoboCup: 10 years of achievements and future challenges. *The AI Magazine*, 28(2):115–132, 2007.

[von Wright, 1963] Georg Henrik von Wright. Practical inference. *The Philosophical Review*, 72:159–179, 1963.

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

[Weirich, 2004] Paul Weirich. *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*. Oxford University Press, Oxford, 2004.

[Williams, 1963] B.A.O. Williams. Imperative inference. *Analysis*, 23, Supplementary 1(3):30–36, 1963.

[Wooldridge, 2000] Michael J. Wooldridge. *Reasoning about Rational Agents*. Cambridge University Press, Cambridge, England, 2000.