Metareasoning as a Formal Computational Problem

Vincent Conitzer

Department of Computer Science
Duke University
Durham, NC 27708, USA

Abstract

Metareasoning research often lays out high-level principles, which are then applied in the context of larger systems. While this approach has proven quite successful, it sometimes obscures how metareasoning can be seen as a crisp computational problem in its own right. This alternative view allows us to apply tools from the theory of algorithms and computational complexity to metareasoning. In this paper, we consider some known results on how variants of the metareasoning problem can be precisely formalized as computational problems, and shown to be computationally hard to solve to optimality. We discuss a variety of techniques for addressing these hardness results.

Introduction

An agent acting in the world generally needs to spend some time and other resources on *deliberation*, to assess the quality of the various plans of action available to it. To find the absolutely optimal plan, an agent generally needs to perform a very large amount of deliberation: it has to consider all the relevant implications of all the relevant facts that it knows about the world, and, if the agent is able to gather additional information, it also has to take all relevant information gathering actions (and consider the implications of the resulting information). This is not always feasible: for example, if the agent has a deadline for choosing an action, there may not be enough time for all of this deliberation. Still, the agent may be able to find a plan of action that is close to optimal. To do so, the agent needs to focus on the parts of the deliberation (*deliberation actions*) that have the greatest impact on the quality of its plan. Determining which deliberation actions to perform is the *metareasoning* problem, in which the agent needs to reason about the reasoning it will perform. Figure 1, taken from Cox and Raja (Cox and Raja 2007), illustrates the three different levels of doing (plans of action), reasoning (deliberation actions), and metareasoning.
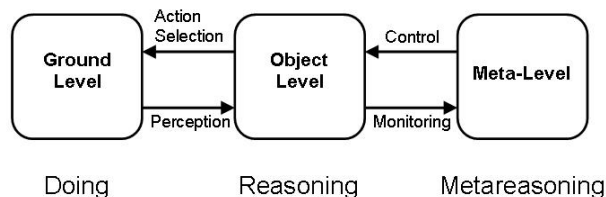
Figure 1: The three levels of doing, reasoning, and metareasoning (Cox and Raja 2007).

While this sounds natural, doing it well is far from an easy problem. For one, the usefulness of one deliberation action may not be seperable from another. For example, if there is a particularly risky plan that the agent is considering, the agent may need to rule out two ways in which this plan could potentially fail. If the agent only manages to rule out one of the two failure possibilities, and does not deliberate on the other, then the plan is still too risky and will not be chosen. Hence, the deliberation on the first failure possibility was a waste of time: the agent does not obtain any benefit from deliberation unless it considers both failure possibilities. To make things more complicated, the agent generally has to consider the outcomes of earlier deliberation actions in choosing the next deliberation action. For instance, in the above example, if the agent considers the first failure possibility and realizes that the plan would in fact fail in this way, then there is no point in considering the other failure possibility, so the agent should spend its valuable time considering other options rather than pointlessly figuring out in exactly how many ways the risky plan would have failed. Hence, in general, the agent does not merely need to choose a subset of the deliberation actions; rather, it needs to create a complete *contingency plan* for deliberating.

From the above, it should be clear that the metareasoning problem is nontrivial, and may in fact be computationally hard. This is an issue of concern, since we

want to avoid the ironic situation in which the agent spends so much time solving the metareasoning problem that there is no time left to take any actual deliberation actions! However, even if the problem does turn out to be computationally hard, this does not mean that we should abandon the metareasoning approach altogether: we could still find fast heuristics or approximation algorithms that find close-to-optimal solutions to the metareasoning problem, or algorithms that find the optimal solution fast under certain conditions.

In the remainder, we first discuss some known results (Conitzer and Sandholm 2003) that imply that certain variants of the the metareasoning problem are in fact computationally hard. While these variants by no means capture all the interesting parts of all metareasoning problems, they are useful for illustrating some computational difficulties that metareasoning systems must face. As such, these results set the stage for the remainder of the paper, in which we discuss what their implications for real metareasoning systems are.

Variants of the metareasoning problem and their complexity

Before we can determine whether the metareasoning problem is computationally hard, we first need to define it as a computational problem. However, there are many different settings in which metareasoning is essential, and each of these settings leads to a different variant of the metareasoning problem. We could try to create a computational definition of the metareasoning problem that is so general that it captures every variant that we might reasonably encounter. It would not be very surprising if such a general problem turned out to be computationally hard; moreover, it is not clear that such a hardness result would tell us anything very interesting, because it could still be the case that most reasonable variants are in fact quite easy to solve.

Instead, we will consider definitions of some very restricted variants of the metareasoning problem that still turn out to be computationally hard. Such results are much more meaningful, because it seems likely that most real-world metareasoning systems need to solve a problem that is at least as hard as at least one of these problems. We discuss mostly the results of Conitzer and Sandholm (Conitzer and Sandholm 2003). Later in the paper, we discuss the implication of these results for the design of metareasoning systems.

Variant 1: Deliberation that leads to predictable improvements

As we mentioned above, one of the main difficulties in metareasoning is that the outcomes of the deliberation actions are uncertain, and what deliberation action should be taken next in general depends on the outcomes of the current and past deliberation actions. Hence, in general, a solution to the metareasoning problem consists of a full contingency plan (at least if we aim to solve the problem to optimality).

In this subsection, however, we consider a simplified variant of the metareasoning problem in which the outcomes of deliberation actions are completely predictable. Specifically, suppose that the agent has m tasks that it needs to complete. For each of the tasks, it has a default plan that has some cost; however, by deliberating on the plan more, the agent can reduce this cost. (For example, Conitzer and Sandholm consider a setting where the agent needs to solve m unrelated vehicle routing problem instances, and it can improve the quality of its solution for each routing problem instance by spending more computation on it—that is, it has an anytime algorithm for the vehicle routing problem.) The agent also has a deadline T by which it needs to finalize all of its plans. Finally, we assume that for each task i, there is a function f_i, where $f_i(t_i)$ is the reduction in the cost of the plan for the ith task that results from spending t_i units of deliberation time on that task. Of course, in reality, this improvement is not so perfectly predictable, but these functions are often used as a modeling simplification. They are called *(deterministic) performance profiles* (Horvitz 1987; Boddy and Dean 1994; Zilberstein and Russell 1996).

The goal is to obtain the maximum total savings given the time limit. That is, we want to choose the times $t_1, \ldots, t_m$ to spend on deliberating on the tasks, with the goal of maximizing $\sum_{i=1}^{m} f_i(t_i)$, under the constraint that $\sum_{i=1}^{m} t_i \leq T$. Conitzer and Sandholm show that (the decision variant of) this problem is NP-complete, even if the f_i are piecewise linear. In contrast, if we require that the f_i are concave, then the problem can be solved in polynomial time (Boddy and Dean 1994). (Similar results based on concavity are common in metareasoning: see, for example, Horvitz (Horvitz 2001).) However, Conitzer and Sandholm argue that the f_i are generally not concave: for example, anytime algorithms generally go through distinct phases, and often the end of one phase does not produce as much improvement as the beginning of the next phase.

It is quite a negative result that even this simple deterministic variant of the metareasoning problem is hard. Still, the implications of this hardness result for metareasoning are limited, because there are variants of the metareasoning problem that do not include the above problem as a subproblem. In a sense, in the above problem, the deliberation actions reveal new plans of action (*e.g.*, vehicle routes). However, there are many metareasoning settings in which the set of available plans is

known from the beginning, and the only purpose of deliberation is to discover which plan is best. If the results of deliberation were perfectly predictable in such a setting, then we would know from the beginning which plan is best, and hence there would be no point in doing any deliberation. That is, the metareasoning problem only makes sense in this context if the outcomes of the deliberation actions are uncertain. This is the topic of the next subsection.

Variant 2: Deliberation to evaluate a fixed set of plans

In this subsection, we consider a different variant of the metareasoning problem. Suppose there is a fixed set of possible plans of action that the agent can choose from. The agent has some *ex ante* expected utility for each plan. The agent can take some deliberation actions on each plan; depending on the outcome of the deliberation action, the expected utility of that plan changes. (For example, Conitzer and Sandholm consider a setting in which a robot must choose a site for digging for precious metals, and before starting to dig, the robot can perform tests (deliberation actions) at each site that will change its beliefs about what metals may be there.) Each deliberation action requires some time (the amount of time is not necessarily the same for each deliberation action), and there is a deadline. Also, we assume that the agent has a probability distribution over how its beliefs about a plan will change upon taking a deliberation action for that plan. (For instance, in the digging example from above, the agent may believe that if it tests at site A, then with probability .6, after the test it will believe that there is a probability of .1 that there is gold at A, and with probability .4, after the test it will believe that there is a probability .2 that there is gold at A. This implies that before the test, it believes that there is a probability of $.6 \cdot .1 + .4 \cdot .2 = .14$ that there is gold at A.) At the end of the deliberation, the agent will choose the plan that currently has the highest expected utility.

The goal here is to find a deliberation strategy that maximizes the expected utility of the agent. Conitzer and Sandholm show that this problem is NP-hard, even if there is at most one deliberation action (with only two outcomes) per plan. They do not prove that the problem is even in NP; it could be that it is, for example, PSPACE-hard. (A closely related class of problems that has been receiving attention more recently is that of "budgeted learning" problems (Madani, Lizotte, and Greiner 2004; Guha and Munagala 2007).)

The problem that we studied in this subsection has the nice property that each deliberation action only affects the agent's beliefs for a single plan, and for each plan there is only a single deliberation action to choose. In the next subsection, we consider a variant of the metareasoning problem without such properties.

Variant 3: Deliberation to disambiguate state

In the final variant of the metareasoning problem that we consider, the agent knows that the world can be in any one of several states. To obtain nonzero utility, the agent needs to determine (by deliberation) the state of the world *with certainty*. If the agent succeeds in determining the state of the world, then the agent's utility depends on which state it is. The agent has a set of available deliberation actions; the outcome of each deliberation action rules out certain states. The outcome of a deliberation action is not deterministic. (For example, Conitzer and Sandholm consider a setting in which a robot is trying to determine the nature of a gap in the floor in front of it. If it cannot determine the nature of the gap with certainty, it should be conservative and turn around, getting utility zero. If it determines the nature of the gap, it may be able to get past the gap and get some utility, depending on what kind of gap it is.) The agent can take only a certain number of deliberation actions (there is a deadline).

Again, the goal for the agent is to find a deliberation strategy that maximizes its expected utility. Conitzer and Sandholm show that this problem is PSPACE-hard, making it the hardest of the metareasoning problems that we have considered (unless the previous problem also turns out to be PSPACE-hard). They also show that the problem remains NP-hard even if, for every state and every deliberation action, there is only a single possible outcome for that deliberation action when the world is in that state (so that the outcomes of deliberation actions are deterministic).

Implications for metareasoning systems

What is the relevance of these complexity results to the design of metareasoning systems? Of course, this depends first of all on whether the problems that we considered are indeed (sub)problems that need to be solved in real metareasoning systems. It seems likely that they are, but we will consider this question in more detail later in this section. For now, let us consider metareasoning systems that indeed need to solve one of the above problems.

We first note that in all of the above problems, the deliberation that the agent can perform is limited by a deadline. So, all the time that the agent spends on the metareasoning problem (deciding what deliberation actions to take) is time that can no longer be spent on actual deliberation. If the metareasoning problem were

solvable to optimality in polynomial time, then perhaps the amount of time spent on the metareasoning problem would always be negligible. But the hardness results discussed above imply that more than polynomial time will be required on at least some instances of the metareasoning problem (unless P=PSPACE, or P=NP for the easier problems). Now, it is certainly possible that these hard instances do not occur in practice very often. For example, we already noted that the first problem that we studied can be solved in polynomial time if the f_i are concave; and it may well be the case that for a particular application, in practice, these functions are in fact always concave. If so, then the hardness result becomes irrelevant (for this particular application). However, we cannot decide whether this is the case if we do not know what instances occur in practice. Moreover, even if the hard metareasoning instances occur only rarely, we would still like to handle them properly. So, it might be the case that in practice most metareasoning instances lie in a class of "easy" instances, and this would mitigate the problem; but it would not eliminate it.

Once the time spent on the metareasoning problem becomes a significant fraction of the total time until the deadline, the agent faces a moving-target problem: the amount of time left for deliberation changes as the metareasoning problem is being solved—but the time left for deliberation is part of the input of the metareasoning problem. This can be resolved in various ways. The simplest is to budget some fixed amount of time B for metareasoning, so that the time for taking deliberation actions is $T - B$, where T was the original deadline. $T - B$ is thus the deadline that is used in the input of the metareasoning problem. This approach requires us to have a metareasoning algorithm that is guaranteed to give a reasonable solution in B units of time; this could be an approximation algorithm with a running time bound of B, or an anytime algorithm that we can simply interrupt at time B.

Another approach would be to have a metareasoning algorithm that, in the first B_1 units of time, finds a solution that requires $T - B_1$ units of deliberation time; then, in the next B_2 units of time, it finds a solution that requires $T - B_1 - B_2$ units of deliberation time; *etc.* At the end of each phase, we can stop this metareasoning algorithm and use the latest solution it provided. One reasonable termination condition for such a metareasoning algorithm is the following: stop when the quality of the solution decreased in the latest phase. One downside to this general approach is that it is not clear that we can re-use any of the computation performed in one phase of the metareasoning algorithm in a later phase, because the phases are effectively solving different problem instances. Still, there seems to be some hope for such re-use, because these instances are closely related. Another key question is how to set the B_i; this may be done dynamically, based on the

solutions found in the earlier phases. This type of analysis of the metareasoning algorithm is getting us into meta-metareasoning.

We have implicitly assumed so far that the metareasoning problem is solved first, and then deliberation starts. This makes sense in a setting where the results of deliberation are deterministic, as in the first problem that we studied. However, if the results of deliberation are not deterministic, then it may make sense to interleave metareasoning and deliberation actions. The results of the deliberation actions will allow us to prune the search space for metareasoning. For instance, in the digging example, if we do the metareasoning first, then we have to consider both the case where the test for gold at site A turns out positive, and the case where it turns out negative. However, if we have already decided that the first deliberation action should be to test for gold at A, then we should go ahead and perform this test before we return to metareasoning: if the test turns out (say) positive, then we no longer need to consider what we would have done if the test had turned out negative.

Finally, let us briefly return to the question of whether our variants of the metareasoning problem were the right ones to study. One debatable aspect is that each of these variants has a deadline that limits the amount of deliberation that can be performed. While having such a deadline is realistic in many situations, we could also consider a model in which there is no deadline, but each deliberation action comes at a cost. Indeed, such models are common in metareasoning, for example in the work on using the expected value of computation to determine when to stop computing (Horvitz, Cooper, and Heckerman 1989; Horvitz and Breese 1990). This modification from a deadline-based model to a cost-based model can affect the complexity of metareasoning. For example, after this modification, the first problem that we studied (where the effect of deliberation is deterministic) becomes easy to solve: now, for each task i *separately*, we can determine the optimal amount of deliberation t_i^*— that is, $t_i^* \in \arg\max_{t_i} f_i(t_i) - ct_i$ (where c is the cost of a unit of deliberation time). Effectively, even though the tasks are *in and of themselves* unrelated, the deadline caused the decisions about how much time to spend on each task to become interrelated; if we switch to the cost model, this effect disappears.

Conclusions

Metareasoning research often lays out high-level principles, which are then applied in the context of larger systems. While this approach has proven quite successful, it sometimes obscures how metareasoning can be seen as a crisp computational problem in its own right. This alternative view allows us to apply tools from the theory

of algorithms and computational complexity to metareasoning. In this paper, we saw how to formulate variants of the metareasoning problem as computational problems, and that these computational problems are generally hard. This approach to the metareasoning problem has at least the following benefits:

- Crisp computational formulations of the metareasoning problem make it easier to consider the key variants of the problem, and to determine what makes the problem hard.
- The hardness results force us to confront the fact that optimal metareasoning is not computationally feasible in general, so that we have to consider approximation algorithms, heuristics, and anytime algorithms for the metareasoning problem, as well as more involved meta-metareasoning approaches.

The reader may be disappointed that we have not given a single, all-encompassing definition of the general metareasoning problem. Certainly, it seems difficult to create such a definition: it seems likely that one would leave out some aspect of the problem. Nevertheless, it may well be interesting to attempt such a definition, even if it is for no other purpose than to provide a starting point for discussion. However, the difficulty of giving a truly general definition is not the main reason that we focused on more restricted variants in this paper. Rather, the main reason for this is that these simple variants are *already* computationally hard. Any fully general definition of the metareasoning problem would presumably include all of these variants as special cases. Because computational problems inherit the hardness of their special cases, that means that we have already shown that the general metareasoning problem is hard (even without giving its precise definition!).

The above also casts some doubt on the value of coming up with a single, general definition of the metareasoning problem as a formal computational problem. If one were to write an algorithm to solve such a general problem, it would have to address a tremendous variety of complexities, including all the ones studied here as well as, presumably, numerous others. As long as we are not trying to solve the general AI problem, it is probably more productive to focus on the special cases of the metareasoning problem that are important for the application at hand, thereby avoiding some of the irrelevant complexities. This is absolutely not to say that there is no value in *studying and discussing* metareasoning in general: in fact, doing so is vital to help us understand the relationships among the different variants of the metareasoning problem, and will allow for the smooth transfer of techniques across these variants.

There are still many open questions. For many key variants, the complexity of the metareasoning problem has not yet been established (including many variants where there is no deadline but deliberation is costly). For the variants where optimal metareasoning has been shown to be hard, there is still a need for approximation algorithms or inapproximability results, as well heuristics and anytime algorithms without formal guarantees but with good practical performance. Perhaps more importantly, it is not yet entirely clear what the best high-level framework is for metareasoning when optimal metareasoning is hard, especially when time is limited and the metareasoning is using up time that could have been used for deliberation actions. As discussed above, we run into the issue that the problem instance becomes a moving target, because the available time for deliberation is changing. Moreover, when the results of deliberation are uncertain, it makes sense to interleave metareasoning and deliberation, because the outcomes of deliberation actions will allow us to prune the possibilities that the metareasoning algorithm no longer needs to consider.

References

Boddy, M., and Dean, T. 1994. Deliberation scheduling for problem solving in time-constrained environments. *Artificial Intelligence* 67:245–285.

Conitzer, V., and Sandholm, T. 2003. Definition and complexity of some basic metareasoning problems. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI)*, 1099–1106.

Cox, M. T., and Raja, A. 2007. Metareasoning: A manifesto. Technical Report BBN TM-2028, BBN Technologies. www.mcox.org/Metareasoning/Manifesto.

Guha, S., and Munagala, K. 2007. Multi-armed bandits with limited exploration. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*.

Horvitz, E. J., and Breese, J. S. 1990. Ideal partition of resources for metareasoning. Technical Report KSL-90-26, Stanford University Computer Science Department.

Horvitz, E. J.; Cooper, G. F.; and Heckerman, D. E. 1989. Reflection and action under scarce resources: Theoretical principles and empirical study. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI)*, 1121–1127.

Horvitz, E. J. 1987. Reasoning about beliefs and actions under computational resource constraints. In *Proceedings of Third Workshop on Uncertainty in Artificial Intelligence*, 429–444. Seattle, Washington: American Association for Artificial Intelligence. Also in L. Kanal, T. Levitt, and J. Lemmer, ed., *Uncertainty in Artificial Intelligence 3,* Elsevier, 1989, pps. 301-324.

Horvitz, E. 2001. Principles and applications of continual computation. *Artificial Intelligence* 126:159–196.

Madani, O.; Lizotte, D. J.; and Greiner, R. 2004.

Active model selection. In *Proceedings of the 20th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 357–365.

Zilberstein, S., and Russell, S. 1996. Optimal composition of real-time systems. *Artificial Intelligence* 82(1–2):181–213.