Decision-Theoretic Deliberation in Resource Bounded Self-Aware Agents

Guido Boella
Dipartimento di Informatica
Università di Torino, Italy
guido@di.unito.it

Joris Hulstijn
Vrije Universiteit, Amsterdam,
jhulstijn@feweb.vu.nl

Leendert van der Torre
SEN-3 - CWI Amsterdam
torre@cwi.nl

Abstract

In this paper we introduce the notion of decision-theoretic deliberation, which studies the relation between classical game and decision theory on the one hand, and agent theories of deliberation on the other hand. We aim at modelling the commonsense notion of intention in systems which are self-aware of their bounded reasoning power concerning their decisions. We propose a transparent decision theoretic deliberation model that contains besides standard actions that change the world and standard actions that only change the information state of the agent (capturing the standard notion of the value of information), also actions that change the agenda of the agent. The agenda contains book keeping of the agent's decision theoretic planner, for which we use the DRIPS planner, and is used to deal with bounded resources, such as the trade-off between further deliberation or starting to execute actions, or whether to change the agenda in case of unexpected observations. This model is the basis of our study of intention. Creating an intention corresponds to putting something on the agenda, and reconsidering intentions corresponds to removing things from the agenda. We show that, in contrast to Cohen and Levesque's approach that defines intention in terms of commitment, our treatment of commitment is based in terms of well understood concepts that consider both creation and reconsideration of intention. Consequently, intention cannot be defined locally in a state of the world, but only globally. Moreover, in contrast to abstract approaches such as Rao and Georgeff's characterization of intention in terms of commitment strategies, we relate intention to planning. Finally, in contrast to Bratman's theory of intention, we do not restrict ourselves to classical planning but we incorporate insights from more recent decision-theoretic planning. We illustrate our decision theoretic deliberation model and the related characterization of intention by a detailed example.

1 Introduction

Agent theory proposes to model the behavior of complex software systems in terms of commonsense mental attitudes like belief, desires, goals, intentions and obligations, ranging from, e.g., the PRS system [Rao and Georgeff, 1992] to the more recent BOID architecture [Broersen *et al.*, 2002] and normative multiagent systems [Boella and van der Torre, 2006]. Decision-theoretic deliberation captures concepts and reasoning mechanisms from agent theory in standard decision-theoretic terms. Thus far, several partial results have been obtained. The relation between beliefs as well as defaults and probabilistic techniques has been studied for some time; there are characterizations of desires and goals in decision-theoretic terms [Herzig *et al.*, 2003]; there are various interpretations of obligations and norms, for example as social laws [Shoham and Tennenholtz, 1997], and there are preliminary results on intention [Boella, 2002b]. See the comparison paper [Dastani *et al.*, 2003] for an overview.

The most problematic issue in decision-theoretic deliberation appears to be the characterization of goal and intention. Roughly, whereas beliefs have been related to probabilities, desires to utilities, and obligations to social laws, goals and intentions do not seem to have an obvious counterpart in classical game or decision theory. Since most discussions on the popular BDI model have focussed on the role of intention in deliberation [Bratman, 1987], we believe that intention is a benchmark example of decision-theoretic deliberation. Goals are more difficult to characterize than intentions for several reasons. First, goals have been studied for a longer time and in more depth, and therefore many different kinds of goals have already been defined and used. Secondly, and more importantly, as observed by [Doyle, 1980], goals have a desirability aspect as well as as intentionality aspect, and therefore have a more complex structure. Goals have been characterized by [Simon, 1955] as utility aspiration levels, a kind of utility threshold, but [Boutilier, 1994] argues that the resulting notion of goal must be generalized, and defines goals in terms of ideality (that express the desirability of states) and the agent's knowledge. Thirdly, goals can be adopted from other agents. Summarizing, the notion of goal is more problematic than is often assumed [Dastani and van der Torre, 2002], but for a characterization of decision-theoretic deliberation it seems better to start with the notion of intention.

In our decision-theoretic model, goals are fixed sets of states build into the underlying decision-theoretic planner.

Intention has been related to choice and commitment [Cohen and Levesque, 1990], where choice can be interpreted as a decision-theoretic notion, but commitment remains as difficult to characterize as intention itself. According to [Bratman, 1987] it is commitment that distinguishes intentions from other motivational attitudes such as goals. The stability of commitment means two things: first, "intentions resist reconsideration and in that sense have inertia" [Bratman, 1987, p.30]. The rationale behind commitment is the resource-boundedness of agents ([Simon, 1955]): agents cannot afford continuous reconsideration and revision of their intentions. Second, it is irrational to maintain an intention without ever reconsidering it when the world changes. Only systems which are self-aware of their bounded reasoning power concerning their decisions are able to trade-off the effort devoted to planning with the need of reconsidering their plans. So the question becomes: under what circumstances does it make sense to reconsider [Schut and Wooldridge, 2001]? For similar reasons, [Zilberstein and Russell, 1993] propose a meta-deliberation approach to reasoning under bounded rationality. In other words, whereas beliefs, desires, and obligations can be characterized by acceptance conditions [van der Torre and Tan, 1999; Veltman, 1996], intentions are typically characterized by their dynamic properties, such as persistence and reconsideration conditions. Finally, theories of intention were developed in the context of resource bounded planning, e.g., [Pollack, 1990], but have not taken advantage of recent developments in decision-theoretic planning and qualitative decision theory.

Our approach to decision-theoretic deliberation – including the characterization of intention – is based on an abstract decision-theoretic model of deliberation in self-aware systems which can express the execution of both normal actions and deliberation actions. We start from [Boella, 2002a; 2002b] on intentions in the context of a decision theoretic planner, called DRIPS [Haddawy and Hanks, 1998]. This work implicitly contains a notion of intention. However, the definition is hidden in the algorithm executed by the planner. To get results concerning intentions, one would have to do extensive simulations, which are difficult to interpret. We therefore present a transparent deliberation model with a decision theoretic interpretation which provides the basis of a study of intentions. We then distinguish between a static and dynamic external world.

Static world. In a static world, in which no unexpected observations occur, the agent only faces a trade off between deliberation and acting. We assume that if the agent deliberates, then it puts plans on the agenda. When the agent starts executing, he chooses randomly a plan that fits the agenda. Putting things on the agenda corresponds to the creation of an intention. Using the DRIPS planner technology, the agent is able to compute upper and lower bounds on expected utility of plans, and therefore can estimate whether further deliberation is worth the effort, given the time discount on its utility function. This captures the resource bounded nature of intentions, but it does not capture the fact that intentions can be reconsidered, or that the agent is committed to its agenda.

Dynamic world. When the world changes, the agent has a trade-off between updating the agenda, or deliberating about a new agenda. We say that an agent is committed to a plan on the agenda, when it does not automatically remove it in case of unexpected changes. Moreover, if an agent is committed to a plan in this sense, and it is put on the agenda as described above in the simpler static world case, then we say that the plan is an intention. Whether a plan is an intention is therefore not a property that is determined only by the point in which the intention is created, but it is also determined by the point in which the world changes and the plan might be dropped from the agenda. Intention is therefore not a local property, but it also has global aspects. In the definition of intention = choice + commitment, this global character is hidden in the definition of commitment. And this is therefore the reason that this definition is non informative.

The layout of the paper is as follows. Section 2 contains the definitions of the deliberation model. Section 3 contains an example to illustrate the use of the definitions. Section 4 discusses in more detail how intention can be analyzed in terms of this model. Related work and conclusions end the paper.

2 Deliberation Model

We first distinguish between the planning environment and the states of the system. The planning environment is everything we consider to be fixed for the planning agent. Inspired by the DRIPS planner, the planning context consists of the propositional variables (X), containing actions or decision variables (A) ordered in an hierarchy (H). Actions without descendants are called atomic, other actions are abstract. Plans are sequences of actions, if they only contain atomic actions then they are called primitive plans, otherwise partial plans. The hierarchy tells us indirectly how partial plans break down in other plans. Moreover, we assume that the roots of the action hierarchy are goals. In other words, the hierarchy only contains compositions of actions for prefixed goals. Finally, using the planner DRIPS [Haddawy and Hanks, 1998] we are able to model partial plans (defined by [Bratman, 1987] as intentions) with associated utility estimates. Two partial functions U^- and U^+ represent the time-discounted utility of the outcome of the best and worst primitive plan subsumed by plan, given a context.

We write A^* for the set of finite sequences built from A, and $Lit(X)$ for the set of literals built from X, and R for the set of real numbers.

Definition 1 (Planning Environment)
A *planning environment* is a tuple $\langle X, A, H, U^-, U^+ \rangle$ where X is a set of propositional variables, $A \subseteq X$ is a set of actions, and $H \subseteq A \times A^*$ is a finite hierarchy that relates each action with a set of sequence of actions, and $U^-, U^+ : Lit(X) \times A^* \rightarrow R$ are partial functions from sets of literals built from X and sequences of actions to real numbers, such that for $s \in Lit(X) \times A^*$, $U^+(s) \geq U^-(s)$. If $(a, (a_1, \ldots, a_n)) \in H$ then doing all of

$a_1, \dots, a_n$ is a way to see to it that a. Since H is a relation, there may be several ways to decompose a. The following notions are defined in a planning environment:

- $AA = \{a \mid \nexists A' \subseteq A, (a, A') \in H\}$. Atomic actions are actions without successors in the hierarchy,
- $BA = A \setminus AA$. Abstract actions are actions that are not atomic,
- $P = A^*$. A plan is a finite sequence of actions.
- $PP = AA^*$. A primitive plan is a plan with no abstract actions,
- $AP = P \setminus PP$. A partial plan is a plan which is not primitive,
- $(a_1, \dots, a_n) \leq (a_1, \dots, a_{i-1}, b_1, \dots, b_m, a_{i+1}, \dots, a_n)$ iff $(a_i, (b_1, \dots, b_m)) \in H$. The ordering on plans is derived from action hierarchy,
- $G = \{a \mid \not\exists a' \in A, A' \subseteq A,$ such that $a \in A'$ and $(a', A') \in H\}$. A goal is an action without a parent in the hierarchy.

A state in a planning environment is composed of an information state I, an agenda Δ containing a set of plans, not of actions. The agenda may reflect the agent's intentions, but it may also contain all other kinds of information. Consider for example an agenda which is kept by a secretary for her boss. The secretary makes all kinds of appointments, and moreover, when there are meetings she writes them down. However, she does not know whether her boss intends to actually go to these meetings. In this paper we assume that the agent will execute a plan in its agenda, but it may be that it drops a plan as soon as something unexpected happens.

Definition 2 (States) Let $\langle X, A, H, U^-, U^+ \rangle$ be a planning environment. A *state* is a tuple $\langle I, \Delta \rangle$, where $I \subseteq Lit(X)$ is a set of literals built from propositional variables X, and the agenda $\Delta \subseteq P$ is a set of plans.

A decision process is like a Markov decision process with deliberation actions which change the state of the agent – in particular its agenda – or execution actions which affect the real world. There are two sources of non-determinacy in our model. First, even in the 'static world' case in which the world and the information state does not change, the agent may execute a random action that is subsumed by its agenda. The reason that we included this option in our model is that our agent can now make a trade-off between deliberating – which leads to a better plan but which also has a cost – and executing actions randomly (given the agenda). The second source of non-determinacy in our model is that the world may change. As a consequence of some actions, the world may change and assuming full observability, also the information state changes. In general rewards can be assigned to states or transitions; here we assign them to states.

Definition 3 (Decision process) A *decision process* is a tuple $\langle S, DA, T, R \rangle$ where S is the set of states defined above, DA a set of deliberation actions, $T : S \times DA \times S \to [0, 1]$ a transition function that associates with every pair of states s, s' and every deliberation action $a \in DA$ a probability in $[0, 1]$ such that for all $s \in S, a \in DA$ we have $\Sigma_{s' \in S} T(s, a, s') = 1$, and R a reward function describing the value assigned to a state.

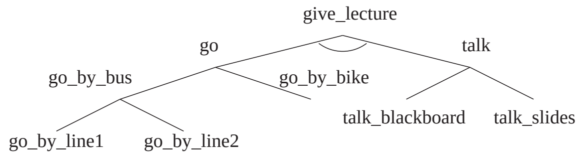

Figure 1: Example action hierarchy.

Our model is a decision-theoretic model of deliberation for the following two reasons. First, we do not model the real world, only the information state of the agent. After an observation, the agent does not know what the result will be. The agent has at most some expectations about the world. Second, we model the agenda as part of the state. One could also model the agenda as part of the information state, but in this way we are able to identify actions which are only concerned with changing the agenda.

Deliberation actions DA have access to the agent's internal state $s \in S$ and modify it by updating its information according to observations and completed actions, and by updating its agenda by refining partial plans and selecting more promising ones. Since the utility functions are partial, it is possible that, to compare plans, the agent must compute the utility of a plan in a new state. For simplicity, we assume that the agenda deals with one plan at a time.

Definition 4 (Deliberation actions) For each $d \in \{refine, execute, revise\}$ define $T(\langle I, \Delta \rangle, d, \langle I', \Delta' \rangle)$, as follows:

Refine $I' = I$, and if $p \in \Delta$ the new $\Delta' = (\Delta \setminus \{p\}) \cup \{p_1, \dots, p_n\}$ where $p \leq p_i$, as defined in terms of H, such that $p' \notin \Delta'$ when $p \leq p'$ and $p' \leq p_i$, and $p'' \notin \Delta'$ when $U^+(p'') < U^-(p_i)$ for $1 \leq i \leq n$.

Execute

If there is only a primitive plan $(a_1, a_2, \dots, a_n) \in \Delta$, information state $I' = I \cup \{a_1\}$ and agenda $\Delta' = (\Delta \setminus \{(a_1, a_2, \dots, a_n)\}) \cup \{(a_2, \dots, a_n)\}$.

If the agenda contains partial plans or more than one primitive plan $p \in \Delta$:

1. Let M be the set of primitive plans subsumed by the current agenda, where $\leq^*$ is the transitive closure of $\leq$:
   $M = \{p' \in PP \mid p \leq^* p'$ for $p \in \Delta\}$.
2. Let E be the set of initial actions of M: $E = \{a_1 \mid (a_1, a_2, \dots, a_n) \in M\}$.
3. Let D_a be the set of sets of primitive plans with the same initial action $a \in E$:
   $D_a = \{\{p_1, \dots p_r\} \mid p_i = (a, a_{i_1}, \dots, a_{i_l})$ and $p_i \in M$ for $1 \leq i \leq r\}$
4. For each $a \in E$, let $I' = I \cup \{a\}$ and let Δ' be the set of plans subsuming the primitive plans whose first action is a:
   $\Delta' = \{p' \mid \{p_1, \dots p_l\} \in D_a$ and $p_i = (a, a_{i_1}, \dots, a_{i_r})$ and
   $p' \leq (a_{i_1}, \dots, a_{i_r})$ and $\not\exists p'' p' \leq p'', p'' \leq (a_{i_1}, \dots, a_{i_r}), 1 \leq i \leq l\}$

5. Suppose S' be the set of states reachable by execution of a single initial action in E; then $|S'| = |E|$ and the probability that action $a \in E$ is executed is $|D_a|/|E|$.

Revise $I' = I$, and if $p \in \Delta$ and $\exists p'$ such that $p' \leq p$ and $\nexists p''$ such that $p'' \leq p$ and $p' \leq p''$ then $\Delta' = (\Delta \setminus \{p\}) \cup \{p'\}$.

In order to construct a policy for the decision process, a reward function R is needed. The reward is based on the agent's internal state, in particular the lower and upper utility bounds of the plans on the agenda. The agent will maximize its utility by executing plans, thus the policy drives the agent's planning activity towards the best solution, taking into account its resource boundedness: refining the current partial solution and computing the utility require time, while the utility of a plan is time discounted; so, the agent has bounded computational resources when it has to build a solution.

Note that the deliberation actions of the decision process consider partial solutions which thus have an uncertain outcome, expressed by the lower and upper bound of a utility function. As suggested in [Boella, 2002b], the reward function does not only consider the action with the least uncertain outcome, but also the ambiguity of the outcome. For this reason the reward function is defined not only as maximizing the outcome but also at minimizing the uncertainty of the outcome by means of refinement of partial plans.

3 Example

The example is based on the following scenario. Our agent has the goal to give a lecture ($give\_lecture$) at the university. But before starting his talk, he must first reach the university. Thus the abstract action $give\_lecture$ corresponds to a plan which contains the abstract actions go followed by $talk$. The agent can reach the university by two alternative ways: going by bus (go_by_bus), with a further choice between bus line 1 and 2 (go_by_line1, go_by_line2), or going by bike (go_by_bike). A talk can be delivered either using the blackboard ($talk\_blackboard$) or using some slides ($talk\_slides$). This example hierarchy is shown in Figure 1. The nodes of the tree are the actions; non-leaves are abstract actions, for example go or $talk$. A sequence of (abstract) actions is a plan. In the example, we denote a plan (a, b) with notation $a; b$. A plan without abstract actions, such as $go\_by\_bike; talk\_slides$, is a primitive plan.

Example 1 (Lecture)

$$\begin{array}{lcl} X & = & A \cup \{rain\} \\ A & = & \{go\_by\_line1, go\_by\_line2, go\_by\_bike, \\ & & talk\_blackboard, talk\_slides, give\_lecture, go, \\ & & talk, go\_by\_bus\} \\ H & = & \{(give\_lecture, (go, talk)), (go, (go\_by\_bus)), \\ & & (go, (go\_by\_bike)), (go\_by\_bus, (go\_by\_line1)), \\ & & (go\_by\_bus, (go\_by\_line2)), \\ & & (talk, (talk\_blackboard)), (talk, (talk\_slides))\} \end{array}$$

With respect to the literals X that make up the information state of an agent, there are two kinds. Information resulting from observations, such as $rain$, or information resulting from completed atomic actions, such as go_by_bike. In Figure 3 and 4 below this kind is marked with a '*'. Alternatively we could record the effects of actions, like $at\_univerity$ as a result of go_by_bike. Currently, the example does not contain conditional actions nor effects.

The refinement structure of partial and primitive plans of the example is depicted in Figure 2. A node corresponds to a set of states, namely those states that have the same partial plan on the agenda. The branches of the tree represent the refinement steps. Please note that, for readability, we assume in this example that the agent always refines the initial abstract action of a plan, first. If we drop this assumption, the state space would be slightly more complicated. The utility boundaries are shown too. In practice, utility boundaries are calculated by the DRIPS planner. Roughly, this works as follows. For end states, we can check if the overall goal G was reached. If so, the state gets a value of 0. If not, the state is not even considered. From this value, we subtract the costs of the actions to get to this state. For non-end states, we take the minimal and maximal utilities of those states that can be reached by a deliberation step from that state. Thus utilities are propagated 'upwards'.

In the scenario, let us suppose that our agent assigns a cost of 0 to using the blackboard. Preparing slides takes an extra effort: -1. Going by bike is for free, 0, but going going by bus requires a fare: -1.5 for line 1, and -2 for line 2.

3.1 Static world

We first consider a scenario where the world is static: apart from the changes due to the execution of actions by the agent, the world remains the same. In particular, what remains the same is the assignment of utilities to the plans the agent is committed to. Consider the following trace, which illustrates that planning need not be completed before the agent can execute actions.

$$\begin{array}{lll} s_1 & \langle\{\}, \{(go, talk)\}\rangle & -\text{refine} \rightarrow \\ s_2 & \langle\{\}, \{(go\_by\_bike, talk)\}\rangle & -\text{execute} \rightarrow \\ s_7 & \langle\{go\_by\_bike\}, \{talk\}\rangle & -\text{refine} \rightarrow \\ s_8 & \langle\{go\_by\_bike\}, \{talk\_blackboard\}\rangle & -\text{execute} \rightarrow \\ s_9 & \langle\{go\_by\_bike, talk\_blackboard\}, \{\}\rangle & \end{array}$$

Note that in state s_2 the plan $go; talk$ has been refined to $go\_by\_bike; talk$ but not to $go\_by\_bus; talk$ even if they are both subsumed by $go; talk$. The reason is that $go\_by\_bus; talk$ with utility [-3,-1.5] is dominated by $go\_by\_bike; talk$ with utility [-1,0]. So $go\_by\_bus; talk$ can be pruned without loss, since all its possible refinements cannot perform better (-1.5) than any refinement of $go\_by\_bike; talk$, which has, at worst, utility -1.

For a primitive plan, there is only one way to execute it: by performing the first action. But if the agent has a partial plan on the agenda, there are several ways to proceed. To execute a partial plan we assume the agent chooses at random one of the primitive plans subsumed by the partial plan and executes the first action. At the same time its agenda is constrained to those partial plans which are compatible with the executed action. Such a 'jump' is a useful option when the agent has no time to deliberate.

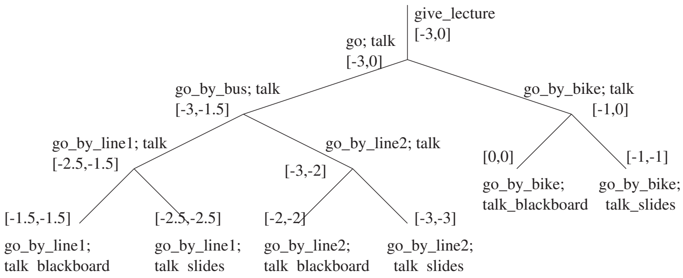

Figure 2: Partial and primitive plans, with utilities.

In the following trace the agent 'randomly selects' an action, and executes it.

$$\begin{array}{ll} s_1 & \langle\{\},\{(go, talk)\}\rangle \qquad -\text{execute} \rightarrow \\ s_7 & \langle\{go\_by\_bike\},\{(talk)\}\rangle \end{array}$$

Since the agent chooses at random, the probabilities depend on the number of primitive plans starting with the same action. Consider for example state s_0 with partial plan $give\_lecture$. If the agent selects to execute, it has to choose at random between the primitive plans subsumed by $give\_lecture$. As is clear from Figure 2 there are 6 of them. Two of these start with go_by_line1, two start with go_by_line2 and two with go_by_bike. So there is a probability of $1/3$ that the agent executes either of these actions. Note that in the resulting state the agent only has the most partial plan compatible with the executed action on his agenda. So in case of go_by_line1 the agenda contains $go\_by\_line1; talk$ rather than $go\_by\_line1; talk\_blackboard$ or $go\_by\_line1, talk\_slides$. Thus the transition probabilities T from state s_0 are as follows, where the states are labelled as in Figure 3.

$$\begin{array}{ll} T(s_0, execute, s_5) = 1/3 & T(s_0, execute, s_6) = 1/3, \\ T(s_0, execute, s_7) = 1/3 & T(s_1, execute, s_5) = 1/3, \\ T(s_1, execute, s_6) = 1/3 & T(s_1, execute, s_7) = 1/3 \\ T(s_2, execute, s_7) = 1 & T(s_3, execute, s_8) = 1 \end{array}$$

In this example we use the deliberation probabilities only for combined refinement and execution of actions; not for refinement itself, nor for revising. For these deliberations, the probabilities are set to 1.

Figure 3 represents the resulting MDP. For reasons of space the name of the actions have been shortened and the name of the deliberation actions abbreviated to E (execution) and R (refinement). Each state is represented by a box containing the name of the state and the commitments of the agents. The utility interval is next to the box. When an action has been completed, it appears labelled with a star. Probabilities are associated to the deliberation transitions.

Up to now we have not introduced the deliberation action *revise*. The reason is that, if the world is not dynamic, it makes no sense to revise ones agenda, since the best strategy does not change: it would be only a waste of resources.

3.2 Dynamic world

Now the example is extended to cover a dynamic world. The MDP is shown in Figure 4. We assume that, according to the agent, the world has some probability to change while the agent is performing a refinement action, since it takes time. If the world changes, primitive plans may not give the same utility. For example, it is possible that it starts raining, with probability 0.1. Moreover, the agent does not want to get wet: utility -5. So the plans that include going by bike now have utility boundaries of [-5,0] or [-6,-5], depending on further costs, like -1 for slides. E.g., in state s_2 the agenda contains $go\_by\_bike; talk$. The agent must still decide whether to talk with or without slides. While the agent is making this decision it may start raining with a probability of 0.1, so that a refinement can lead to two different states. With probability 0.9 it leads to $s_3 = \langle\{\},\{go\_by\_bike; talk\_blackboard\}\rangle$, and with probability 0.1 to state $s_{24} = \langle\{rain\},\{go\_by\_bike; talk\_blackboard\}\rangle$. Even though in both states the agent has $go\_by\_bike; talk\_blackboard$ on its agenda, the utility is different: $[0, 0]$ for s_3 and $[-5, -5]$ for s_{24}.

By contrast, we assume that while the agent is executing an action or revising its agenda the world will not change in a way that affects the utility boundaries. This assumption is sufficient for studying the phenomena we are interested in. Thus, the refinement transitions from s_0 are now changed:

$$\begin{array}{ll} T(s_0, refine, s_1) = .9 & T(s_0, refine, s_{22}) = .1 \\ T(s_1, refine, s_2) = .9, & T(s_1, refine, s_{23}) = .1 \\ T(s_2, refine, s_3) = .9 & T(s_2, refine, s_{24}) = .1 \\ T(s_5, refine, s_4) = 1 & T(s_7, refine, s_8) = .9 \\ T(s_7, refine, s_{25}) = .1 & \end{array}$$

For states which do not involve going by bike, e.g., s_5, the fact that the world changes is not shown, since the rain does not alter their utility.

The consequence of the introduction of a dynamic world is that revision becomes meaningful. Revision (or expansion) is denoted by X in Figure 4. By revision we mean that the agent puts the least partial plan on the agenda that is subsuming the current agenda: if $\Delta = \{p\}$ and $\exists p'$ such that $p' \leq p$ and $\nexists p''$ such that $p'' \leq p \in H$ and $p' \leq p''$ then the revised agenda is $\Delta' = \{p'\}$. For example, in state s_2 the agent has $go\_by\_bike; talk$ on its agenda, while after the revision, in state s_1 the agenda contains $go; talk$ which subsumes the former. Therefore the utility bounds of $go; talk$ include the utility bounds of $go\_by\_bike; talk$.

A revision action is not always the inverse of a refinement, as for example between states s_5 and s_4. The reason is that if the world changes, the utility assigned to plans can change

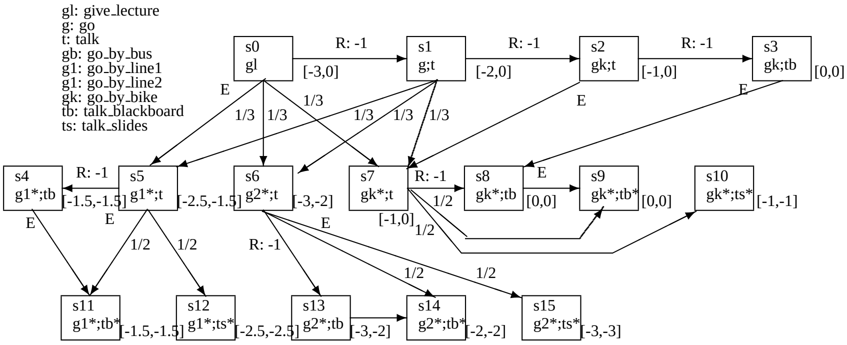

Figure 3: The MDP in a static world.

too. Consider the refinement of state s_1. There are two possibilities: s_2 with probability 0.9 and s_{23} with probability 0.1. While revising from $go\_by\_bike; talk$ to $go; talk$ in state s_2 leads to state s_1 again, revising the same agenda in state s_{23} leads to state s_{22}: since the utility of $go; talk$ subsumes the utility of the plans of going by bike, if their utility changes, the utility of $go; talk$ changes too. It is now $[-6, -1.5]$, and not $[-3, 0]$ like in state s_1.

4 Discussion

In this paper, rather than specifying an algorithm for meta-deliberation as in [Zilberstein and Russell, 1993], we represent meta-deliberation as a decision process driven by the value of information provided by refining partial plans.

When we consider concepts from cognitive science like beliefs, desires, goals, norms, intentions, et cetera, we do not consider a single state, but we consider the whole MDP. Local beliefs are represented by the information states, but more generally the beliefs of the agent are reflected by the transition function of the MDP. For example, the probabilities associated with transitions are part of the beliefs of the agent too. Likewise, desires are associated with utilities, but also with the reward function of the MDP. In this paper we do not consider goals or norms. Whether something is an intention is when it gets put on the agenda, and stays there when things change (depending on type of commitment strategy). We thus consider two issues, the acceptance condition and the reconsideration condition. This can be paraphrased as follows:

intention = accepted and persists

In other words, instead of saying that intentions are choice and commitment, we define it directly in terms of agenda. Moreover, our decision theoretic reconstruction makes this more precise, as follows.

Static world. The discussion in Section 3.1 tells us something about acceptance of intention. A minimal condition of intention is that there is a choice between putting it on the agenda or not, and there is a preference for putting it on the agenda. In our case, the alternative is to execute a plan randomly, and there is a preference for refining a plan when the expected payoff of the new plan pays for the costs of refining the plan.

Dynamic world Moreover, Section 3.2. tells us something about termination or reconsideration. When the information state changes due to an action, such as in state s_{23}, then the agent can choose either to drop its agenda, or to stick to it. For a plan to be called an intention, the agent has to usually stick to its plan. In our decision-theoretic model, it depends on the costs of refinement, and the expected benefits.

The notion of intention can be made more precise by defining a measure that tells us when an agent often keeps his plans on the agenda. If we increase the cost of deliberation, then there will always be a point after which the agent does not deliberate but sticks to its plans. In other words, there is always a cost threshold above which intentions are used into the agent's deliberation.

5 Related work

Several authors addressed the problem of decision making under bounded rationality. The most prominent approaches are the following.

[Boddy and Dean, 1994] proposed *continuous deliberation scheduling*, a planning algorithm based on the idea that an agent has a fixed set of decision procedures to react to events happening in the environment. The quality of the solution of a decision procedure depends on the time given to the procedure. Continuous deliberation scheduling is an algorithm that schedules decision procedures to achieve the highest overall satisfaction.

[Russell and Wefald, 1991]'s *discrete deliberation scheduling* algorithm is based on the idea that at every moment in time an agent must deliberate or act. Discrete deliberation scheduling is an algorithm that decides on the basis of the expected values of deliberation or action, whether to deliberate or to act respectively.

[Russell *et al.*, 1993] propose the notion of bounded optimality: a perfectly rational agent will base its reasoning on decision theory, given what it knows of the environment. In practice, where agents reason in real-time, this type of rationality is not feasible, thus we have to select a subset of these perfectly rational agents that are able to reason in real time. These agents are called bounded optimal agents, and behave as well as possible given their computational resources.

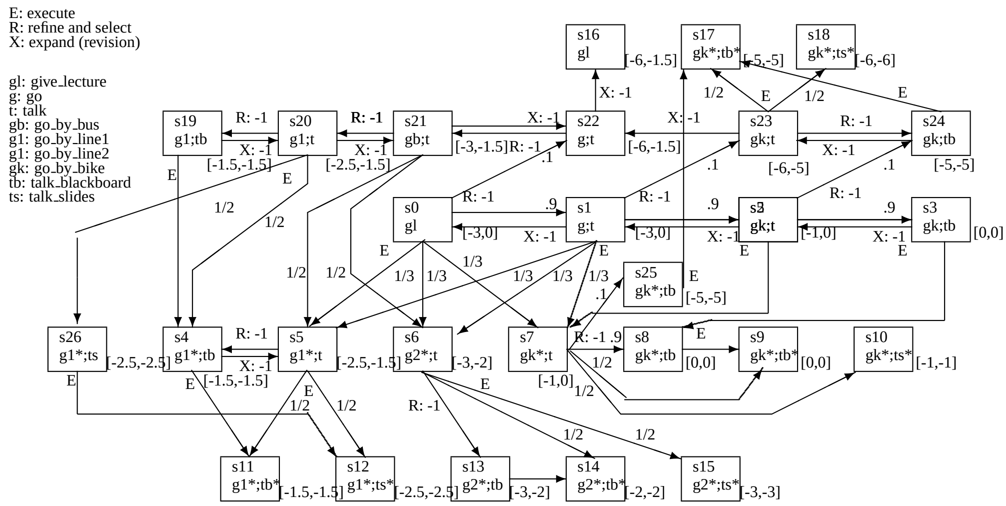

Figure 4: The MDP in a dynamic world.

In studying resource bounded reasoning, [Zilberstein, 1996] proposes to consider planning as a source of information for the execution architecture, at the same level as sensing. In this paper we are inspired by this view of planning, even if we adopted a formalism similar to the DRIPS planner. Like we do, Zilberstein uses a planning anytime algorithm which improves the quality of its solutions as a function of time. Using [Zilberstein and Russell, 1996]'s classification, we consider the dimension of certainty for measuring the quality of a plan. A conditional performance profile makes it possible to predict the advantage of planning a refined solution with respect to the current approximate one. In our framework this performance profile can be reconstructed starting from the states of the MDP.

Some more recent proposals address the problem of the relation of bounded rationality with the BDI model. [Schut and Wooldridge, 2001] study how to model intention reconsideration in belief-desire-intention (BDI) agents. Since their work is based on [Russell and Wefald, 1991], they assume that at any moment in time the agent has some default action it can perform. The agent can either execute this action or deliberate, where deliberation can lead to a better action. By contrast, in our work, we do not simply consider a default action to be executed as an alternative to further deliberation. We allow an agent to execute a random plan which is subsumed by its current partial intention. In this way we model a choice which has no computational costs, unlike for example selecting an optimal plan.

[Schut *et al.*, 2002] study the same problem by using the theory of Markov decision processes (MDP) for planning in partially observable stochastic domains. They view an intention reconsideration strategy as a policy in a partially observable Markov decision process (POMDP): solving the POMDP thus means finding an optimal intention reconsideration strategy. Like in our work they represent both the execution of normal actions and redeliberation actions in the transitions of the MDP, and they build the optimal policy to be used runtime, beforehand. The reward of executing an action is the utility achieved by the action, while the utility of redeliberation is indirectly defined "as the expected worth of future states in which the agent has correct intentions". As intentions resist reconsideration, the implementation of the reward structure should thus favor action over deliberation.

Even if our research goal is similar, there are some important differences with our work. First of all, we do not focus on intention reconsideration, but we want to explain the very notion of intentions by studying a rationally bounded agent model. Second, in their work "the BDI agent can be seen as a domain dependent object level reasoner, concerned directly with performing the best action for each possible situation; the POMDP framework is then used as a domain independent meta level reasoning component, which lets the agent reconsider its intentions effectively". By contrast, we do not have a BDI agent model which is used as a deliberation component; rather we translate the basic steps of an algorithm for BDI agents as deliberation actions which determine the transitions in the MDP. In this paper, for simplicity, we focused on the deliberation actions of refining a partial plan, revising the current plan by making it more partial and executing primitive or partial plans. In particular, we do not assume that a deliberation action is able to lead the agent to a state where it intends a new primitive plan. We provide a more fine-grained model where each step of the agent deliberation is mapped onto a transition of the MDP. Third, since according to [Bratman, 1987] the notion of intention is related to partial plans, we adopted a hierarchical plan formalism inspired on the DRIPS planner. In this way we are able to study the bounded rationality of agents not only in terms of costs of deliberation actions but also in terms of the value of information provided by refining partial plans.

6 Conclusions

The following conclusions can be drawn. Intentions can be analysed as a side-effect of meta-reasoning on action and deliberation in systems which are self-aware of their bounded reasoning power concerning their decisions. In general, devising a new plan has to overcome overhead costs. Continuing the old plan is therefore more beneficial, when the initial utility differences are not very large. The stability, observed as the major function of intentions, therefore automatically comes out of a standard decision theoretic planning setting. Intention gets essentially a dynamic semantics: a semantics in terms of the changes to utilities.

Further research includes several issues. First of all, we are introducing explicit observation actions which are executed if the information they are likely to convey is worth the costs of those actions. Second, we can introduce a more fine grained model of the deliberation process of an agent, e.g., by considering also a deliberation action of evaluating the utility of a (partial) plan, with an additional deliberation cost. Third, we are studying how different probability distributions and different costs of redeliberation strategies affect the persistence of intentions. To aid in the experiments, a prototype implementation in Prolog is being constructed. This will also allow experiments with more realistic examples.

References

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

[Boella and van der Torre, 2006] G. Boella and L. van der Torre. A game theoretic approach to contracts in multiagent systems. *IEEE Transactions on Systems, Man and Cybernetics - Part C*, 2006.

[Boella, 2002a] G. Boella. Decision theoretic planning and the bounded rationality of BDI agents. In *Procs. of GTDT 2002 Workshop. Technical report WS-02-06*, pages 1–10, 2002. AAAI Press.

[Boella, 2002b] G. Boella. Intentions: choice first, commitment follows. In *Procs. of AAMAS'02*, pages 1165 – 1166. ACM Press, 2002.

[Boutilier, 1994] C. Boutilier. Towards a logic for qualitative decision theory. In *Procs. of KR'94*, pages 75–86. Morgan Kaufmann, 1994.

[Bratman, 1987] M. E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge (MA), 1987.

[Broersen *et al.*, 2002] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly*, 2(3-4):428–447, 2002.

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

[Dastani and van der Torre, 2002] M. Dastani and L. van der Torre. What is a normative goal? In *Procs. of RASTA'02 Workshop*, LNCS 2934, pages 210–227, Springer Verlag, Berlin, 2002.

[Dastani *et al.*, 2003] M. Dastani, J. Hulstijn, and L. van der Torre. How to decide what to do? *European Journal of Operational Research*, 2003.

[Doyle, 1980] J. Doyle. A model for deliberation, action and introspection. Technical Report AI-TR-581, MIT AI Laboratory, 1980.

[Haddawy and Hanks, 1998] P. Haddawy and S. Hanks. Utility models for goal-directed decisiontheoretic planners. *Computational Intelligence*, 14(3):392–429, 1998.

[Herzig *et al.*, 2003] Andreas Herzig, Jrme Lang, and Pierre Marquis. Action representation and partially observable planning in epistemic logic. In *Procs. of IJCAI'03*, pages 1067–1072. Morgan Kaufmann, 2003.

[Pollack, 1990] M. E. Pollack. The uses of plans. *Artificial Intelligence*, 57(1):43–69, 1990.

[Rao and Georgeff, 1992] A. Rao and M. P. Georgeff. An abstract architecture for rational agents. In *Procs. of KR'92*, pages 439–449, 1992.

[Russell and Wefald, 1991] S. Russell and E. Wefald. Principles of metareasoning. *Artificial Intelligence*, 49:361–395, 1991.

[Russell *et al.*, 1993] Stuart J. Russell, Devika Subramanian, and Ronald Parr. Provably bounded optimal agents. In *Procs. of IJCAI'93*, pages 338–345. Morgan Kaufmann, 1993.

[Schut and Wooldridge, 2001] M.C. Schut and M. Wooldridge. Principles of intention reconsideration. In *Procs. of AA'2001*. ACM Press, 2001.

[Schut *et al.*, 2002] Martijn C. Schut, Michael Wooldridge, and Simon Parsons. On partially observable MDPs and BDI models. In *Foundations and Applications of Multi-Agent Systems, selected papers of UKMAS'02*, LNCS 2403, pages 243–260. Springer-Verlag, Berlin, 2002.

[Shoham and Tennenholtz, 1997] Y. Shoham and M. Tennenholtz. On the emergence of social conventions: Modeling, analysis and simulations. *Artificial Intelligence*, 94(1–2):139–166, 1997.

[Simon, 1955] Herbert A. Simon. A behavioral model of rational choice. *Qualterly Journal of Economics*, 69:99–118, 1955.

[van der Torre and Tan, 1999] L. van der Torre and Y. Tan. Rights, duties and commitments between agents. In *Procs. of IJCAI'99*, pages 1239–1244, 1999.

[Veltman, 1996] Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.

[Zilberstein and Russell, 1993] S. Zilberstein and S. J. Russell. Anytime sensing, planning and action: A practical model for robot control. In *Procs. of IJCAI'93*, pages 1402–1407, Chambery, France, 1993.

[Zilberstein and Russell, 1996] S. Zilberstein and S. Russell. Optimal compositionof real-time systems. *Artificial Intelligence*, 82:181–213, 1996.

[Zilberstein, 1996] S. Zilberstein. Resource-bounded sensing and planning in autonomous systems. *Autonomous Robots*, 3:31–48, 1996.